PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/705, 16/28, C12N 5/10, C12Q 1/68, G01N 33/68

A2

(11) International Publication Number: WO 95/22609

(43) International Publication Date: 24 August 1995 (24.08.95)

(21) International Application Number: PCT/GB95/00356

(22) International Filing Date: 21 February 1995 (21.02.95)

(30) Priority Data:

9403285.1	21 February 1994 (21.02.94)	GB
9415532.2	1 August 1994 (01.08.94)	GB

(81) Designated States: JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(71) Applicant *(for all designated States except US)*: THE WELLCOME FOUNDATION LIMITED [GB/GB]; Unicorn House, 160 Euston Road, London NW1 2BP (GB).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: MAKOFF, Andrew, Joseph [GB/GB]; Langley Court, Beckenham, Kent BR3 3BS (GB).

(74) Agent: STOTT, Michael, John; The Wellcome Foundation Limited, Langley Court, Beckenham, Kent BR3 3BS (GB).

(54) Title: HUMAN GLUTAMATE RECEPTOR PROTEINS

(57) Abstract

The invention provides human metabotropic glutamate receptors 3 and 4 (human mGluR3 and human mGluR4), DNAs encoding the receptors and a method for screening for compounds which bind to or modulate the activity of the receptors.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

HUMAN GLUTAMATE RECEPTOR PROTEINS

The invention relates to the novel human metabotropic glutamate receptors 3 and 4 (human mGluR3 and human mGluR4) and polypeptide fragments thereof, DNA coding therefor and oligonucleotide fragments thereof, antibodies specific for the receptors and a method for screening for compounds capable of binding to or modulating the activity of the receptors.

In the mammalian central nervous system, L-glutamate serves as a major excitatory neurotransmitter. The interaction of glutamate with its membrane bound receptors is believed to play a role in many important neuronal processes including fast synaptic transmission, synaptic plasticity, and long term potentiation. These processes are fundamental to the maintenance of life and normal human abilities such as learning and memory (Monaghan, D.T. et al., 8 Neuron 267 (1992)).

In addition to its role in normal human physiology, interaction of L-glutamate with its receptors is believed to play a key role in many neurological disorders such as stroke, epilepsy, and head trauma, as well as neurodegenerative processes such as Alzheimer's disease (Olney, R.W., 17 Drug Dev.Res. 299 (1989)). For this reason, understanding the molecular structure of human L-glutamate receptors is important for understanding these disease processes as well as for searching for effective therapeutic agents. To date, the search for therapeutic agents which will bind and modulate the function of human glutamate receptors has been hampered by the unavailability of homogeneous sources of receptors. The brain tissues commonly used by pharmacologists are derived from experimental animals (non-human) and furthermore contain mixtures of various types of glutamate receptors.

In searching for drugs for human therapy, it is desirable to use receptors which are more analogous to those in the intact human brain than are the rodent receptors employed to date.

The discovery of human glutamate receptors, therefore, provides a necessary research tool for the development of selective pharmaceutical agents.

Pharmacological characterisation of receptors for L-glutamate has led to their classification into two families based on their biological function: the ionotropic receptors which are directly coupled to cation channels in the cell membrane, and the metabotropic receptors which function through coupling to G-proteins. To date seven different rat mGluRs and three different human mGluRs have been cloned. These are: rat mGluR1, (Masu et al, Nature, Vol 349, 760 - 765, (1991) , Houamed, K.M., et al. Science, Vol 252, 1318-1321, (1991)); rat mGluR2, (Tanabe, Y. et al., Neuron 8, 166-179, (1992)); rat mGluR3, ibid.; rat mGluR4, ibid.; rat mGluR5, (Abe, T. et al., J. Biol. Chem., 267, 13361-13368, (1992)); rat mGluR6, (Nakajima, Y. et al., J. Biol. Chem., 268, 11868-11873, (1993)); rat mGluR7 (Okamoto, N et al, J. Biol. Chem. vol 269, 1231-1236 (1994), Saugstad, J.A. et al, Mol. Pharmacol vol 45, 367-372 (1994)); human mGluR1 (WO 94/29449 and EP-A-569 240); human mGluR2 (WO 94/29449); human mGluR3 (WO 94/29449) and human mGluR5 (Minakami, R. et al Biochem. Biophys. Res. Comm. vol 199, 1136-1143 (1994) and WO 94/29449). Types 1 and 5 are coupled to phosphoinositol metabolism via phospholipase C. The other five mGluRs are all negatively coupled to adenylate cyclase (Nakanishi, S., Science 258, 597-603, (1992); Nakajima, Y., Iwakabe, H., Akazawa, C., Nawa, H., Shigemoto, R., Mizuno, N. and Nakanishi, S., J. Biol. Chem. 268, 11868-11873 (1993); Okamoto, N. et al, J. Biol. Chem. vol 269, 1231-1236 (1994); Saugstad, J.A. et al, Mol. Pharmacol. vol 45, 367-372 (1994)). Because the mGluRs have only been recently described, they are poorly-defined pharmacologically. However, they appear to fall into three groups by their agonist selectivity: mGluR1 and 5 (quisqualate), mGluR2 and 3 [(2S,1'R,2'R,3'R)2-(2,3-dicarboxycyclopropyl)glycine, DCG-IV] and mGluR4, 6 and 7 (2-amino-4-phosphonobutyrate, AP4) (Nakanishi, S., Science 258, 597-603, (1992); Nakajima, Y.,

Iwakabe, H., Akazawa, C., Nawa, H., Shigemoto, R., Mizuno, N. and Nakanishi, S., J. Biol. Chem. 268, 11868-11873. (1993), Hayashi, Y., Momiyama, A., Takahashi, T., Ohishi, H., Ogawa-Meguro, R., Shigemoto, R., Mizuno, N. and Nakanishi, S., Nature 366, 687-690 (1993), Okamoto, N. et al, J. Biol. Chem. vol 269, 1231-1236 (1994), Saugstad, J.A. et al, Mol. Pharmacol. vol 45, 367-372 (1994)). These groups also reflect the relative sequence homologies of each of the receptors.

The present invention provides a DNA isolate encoding human mGluR3 or human mGluR4. The DNA isolate may encode human mGluR3 having an amino acid sequence at least 97% (e.g. at least 98% or at least 99%) identical to the sequence of SEQ ID NO:2. The DNA isolate may alternatively encode human mGluR4 having an amino acid sequence at least 97% (e.g. at least 98% or at least 99%) identical to the sequence of SEQ ID NO:4. Preferably, the DNA isolate encodes human mGluR3 having exactly the amino acid sequence of SEQ ID NO:2 or human mGluR4 having exactly the amino acid sequence of SEQ ID NO:4.

The nucleotide sequence of the DNA isolate may be at least 90% identical (e.g. at least 95%, at least 96%, at least 97% at least 98% or at least 99% identical) to the human mGluR3 coding sequence set forth from position 253 or 259 to position 2889 of SEQ ID NO:1. The nucleotide sequence may alternatively be at least 90% identical (e.g. at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical) to the human mGluR4 coding sequence set forth from position 171 to position 2906 of SEQ ID NO:3.

In general, the nucleotide sequence of the DNA isolate is exactly the sequence from position 253 or 259 to position 2889 of SEQ ID NO:1 or exactly the nucleotide sequence from position 171 to position 2906 of SEQ ID NO:3. However, these sequences from SEQ ID NOs:3 and 4 may be changed by nucleotide substitution, deletion, insertion or extension. Such changed sequences should still encode a protein which retains mGluR3 or

mGluR4 activity, e.g. L-glutamate binding and/or negative coupling to adenylate cyclase. A substitution, deletion or insertion may involve one or more nucleotides, typically from one to five, one to ten or one to twenty nucleotides.

The human mGluR3 nucleotide and amino acid sequences set forth in present SEQ ID NO:1 contain three differences compared to the human mGluR3 sequences set forth in WO 94/29449 (see SEQ ID NO:5), as follows:

1. There are two internal nucleotide change differences which both result in amino acid changes. Amino acid Gly 334 (GGC) in the WO 94/29449 sequence is Asp (GAC) in the present sequence and Glu 374 (GAA) in the WO 94/29449 sequence is Asp (GAC) in the present sequence.

2. The mGluR3 nucleotide sequence contains two ATG codons in the same reading frame at the beginning of the translated sequence. In WO 94/29449, the earlier of the two codons is assumed to be the start codon so the translated sequence begins MetLysMet.... However, the sequence surrounding the second ATG codon is much more favourable for initiation of translation and it is therefore likely that this codon initiates translation giving a translated sequence which begins MetLeuThr.... It is possible that two protein products are produced from the two initiation codons and both products are included in the invention.

3. The final 17 nucleotides of the 3' untranslated end of the WO 94/29449 sequence differ completely from the corresponding region of present SEQ ID NO:1.

The invention includes an oligonucleotide fragment having a sequence of a portion of a DNA isolate encoding human mGluR3 or mGluR4 of the invention. The fragment may have a sequence at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to sequence encoding human mGluR3

or human mGluR4 of the invention, for example sequence shown in SEQ ID NO:1 or SEQ ID NO:3. The fragment is at least 12 nucleotides in length, e.g. at least 15, at least 18, at least 30, at least 60, at least 180, at least 360 or at least 720 nucleotides in length. The fragment may be single or double stranded. When the fragment is single stranded, it may have a sequence from either a sense or antisense strand of a DNA molecule encoding a human mGluR3 or a human mGluR4 of the invention. An antisense fragment may be useful in the therapeutic treatment of a disease involving over-expression of human mGluR3 or human mGluR4.

The present invention also provides a human mGluR3 protein and a human mGluR4 protein, such as those having the amino acid sequences of SEQ ID NOs:2 and 4. Alternatively, there is provided a human mGluR3 or mGluR4 protein having a sequence which is at least 96% identical to the amino acid sequence of SEQ ID NO:2 or 4, for example a sequence at least 97%, at least 98% or at least 99% identical to the amino acid sequence of SEQ ID NO: 2 or 4.

The invention further includes a polypeptide fragment of a human mGluR3 or mGluR4 protein of the invention, which fragment is at least 8 amino acids in length, for example at least 12, at least 24, at least 48 or at least 96 amino acids in length.

The human mGluR3 or mGluR4 protein will usually be obtained by recombinant DNA techniques. However, the protein may be obtained using biochemical purification of the protein from its natural origin.

The human mGluR3 or mGluR4 is generally expressed on the surface of a host cell, but may be in purified or isolated form. Preferably, the mGluR3 or mGluR4 in purified form comprises a preparation in which at least 80%, at least 85%, at least 90%, at least 95%, at least 98% or at least 99% of the weight of protein in the preparation is human mGluR3 or mGluR4.

According to a further embodiment, there is also provided a DNA isolate encoding a fragment of the amino acid sequence of the human mGluR3 or mGluR4 protein. DNA sequences encoding fragments of the human mGluR3 or mGluR4 protein preferably encode those parts of the amino acid sequence which characterise the receptor, i.e. those parts which are most distinct from other human mGlu receptor proteins.

A person of ordinary skill in the art would by reference to the sequences disclosed herein know how to obtain a DNA isolate according to the invention using methodologies and techniques well known to the skilled person. The various means include for example, DNA synthesis, or more preferably, recombinant DNA techniques. Techniques for constructing recombinant isolates are described by Sambrook et al, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989).

Human mGluR3 or mGluR4 protein and fragments thereof may be obtained by expression of DNA isolates of the present invention in appropriate expression vectors. A library of replicable expression vectors may be created by cloning genomic DNA or, more preferably, cDNA into a parent vector. The library is screened for clones containing the desired nucleic acid sequence, e.g. by means of a DNA probe. Probes having consensus sequences for specific regions of various human mGluR proteins may be used. For example, the following consensus primers designed from the published sequences of rat mGluR1 to 5 proteins may be used:

ACAACGACACACCCGTGGTCAA (SEQ ID NO:5);
TCCGGTCGGGAGCTCTGCTA (SEQ ID NO:6);
GCTACTCTGCCCTGCTGACCAAGAC (SEQ ID NO:7);
GCAATGCGGTTGGTCTTGGT (SEQ ID NO:8);
AAGTTCATCGGCTTCACCATGTACAC (SEQ ID NO:9);
GATGCAGGTGGTGTACATGGTGAA (SEQ ID NO:10); and
TCTCCGGTTGGAAGAGGATGATGT (SEQ ID NO:11).

A replicable expression vector is a vector which contains the appropriate origin of replication sequence for replication of the vector and the appropriate sequences for expression of the foreign nucleotide sequence in the vector. The sequences for expression of a foreign sequence will generally include a transcription promoter operably linked to the foreign sequence. The term "operably linked" refers to a linkage in which the promoter and foreign sequence are connected in such a way to permit expression of the foreign sequence. The transcription promoter sequence may be part of the parent vector sequence into which the foreign sequence is inserted. Alternatively, the promoter sequence may be a native promoter sequence of a gene encoding human mGluR3 or mGluR4 of the invention. A vector may be, for example, a plasmid, virus or phage vector. A vector may contain one or more selectable markers, for example an ampicillin resistance gene in the case of a bacterial vector or a neomycin resistance gene in the case of a mammalian vector. A foreign gene sequence inserted into a vector may be transcribed in vitro or the vector may be used to transform or transfect a host cell.

According to a further aspect of the invention, there is provided a host cell transformed or transfected with a vector into which there has been inserted a DNA isolate (or a fragment thereof) encoding human mGluR3 or mGluR4 protein. A vector and host cell will be chosen so as to be compatible with each other, and may be prokaryotic or eukaryotic. A prokaryotic host may, for example, be E. coli, in which case the vector may, for example, be a bacterial plasmid or a phage vector. A eukaryotic host may, for example, be a yeast (e.g. S.cerevisiae) cell, a chinese hamster ovary (CHO) cell, BHK cell, oocyte or an insect cell (e.g. Spodoptera frugiperda). When the host is an insect cell, the vector is generally a baculovirus vector (reviewed by Luckow and Summers in BIO/TECHNOLOGY, Vol. 6, 47-55 (1988)). When the host is an oocyte mRNA encoding human mGluR3 or mGluR4 protein can be injected into the oocyte for expression of the protein.

A host cell expressing human mGluR3 or a human mGluR4 according to the invention may be produced by a method comprising

(a) transforming or transfecting cells with a vector comprising a DNA isolate (or fragment thereof) encoding human mGluR3 or mGluR4;

(b) culturing the cells under conditions in which the mGluR3 or mGluR4 is expressed; and

(c) recovering a host cell expressing the mGluR3 or mGluR4 from the culture.

An oligonucleotide fragment according to the invention will generally be DNA, although other types of nucleic acid may be used, for example RNA or modified DNA. A number of different types of nucleic acid modification are known in the art. These include methylphosphonate and phosphorothioate backbones, and addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule.

Oligonucleotide fragments may be useful in the detection of nucleic acid sequences. An oligonucleotide fragment corresponding to a portion of the human mGluR3 or mGluR4 coding sequence may be an oligonucleotide probe or an oligonucleotide primer (e.g. a polymerase chain reaction (PCR) primer), which will hybridise to a nucleic acid molecule (e.g. a DNA or RNA molecule) encoding a human mGluR3 or mGluR4 protein of the invention. The probe or a pair of primers may be used to detect or quantitatively determine the nucleic acid sequence. This has diagnostic utility in detecting and quantitatively determining human mGluR3 or mGluR4 mRNA associated with a disease state, for example disorders of the central nervous system.

A fragment which is a probe or primer may carry a revealing label, such as 32P, digoxigenin or biotin. Preferably, the probe or primer will specifically hybridise only to its target sequence, e.g. a portion of the sequence disclosed in SEQ ID

NO:1 or SEQ ID NO:3, and not to other sequences. A probe or primer which hybridises only to its target sequence will generally be exactly complementary to the target sequence whereas a probe or primer which is only selective may have a sequence which is, for example, at least 80%, at least 85%, at least 90%, at least 95%, at least 98% or at least 99% complementary to the target sequence. A probe which is not exactly complementary to its target sequence has utility in the identification of new human mGluR coding nucleotide sequences. A fragment which is a probe or primer may have from 12 to 60 nucleotides, e.g. from 12 to 40 nucleotides, or from 15 to 30 nucleotides.

Primers for PCR are generally provided as a pair. A first primer hybridises to a sense sequence 3′ to the sequence to be amplified and a second primer hybridises to an antisense sequence 5′ to the sequence to be amplified. This allows synthesis of double stranded DNA representing the region between the two primers.

Thus, the invention provides a method of amplifying a target nucleic acid sequence present in a nucleic acid encoding a human mGluR3 or mGluR4 protein of the invention, which method comprises carrying out PCR employing a primer of the invention. Such a method generally comprises carrying out cycles of

(a) denaturing double stranded DNA containing the target sequence to obtain single stranded DNA;
(b) hybridizing a first primer to a sense strand 3′ to the target sequence, and hybridising a second primer to an antisense strand 5′ to the target sequence; and
(c) synthesising DNA from the first and second primers.

The number of cycles is suitably from 10 to 50, preferably 20 to 40, more preferably 25 to 35. The method may be carried out starting from a double stranded nucleic acid (e.g. dsDNA) or a single stranded nucleic acid (e.g. mRNA). The target sequence

may be a complete human mGluR3 or mGluR4 protein encoding sequence or a partial human mGluR3 or mGluR4 protein encoding sequence.

As will be appreciated by a person skilled in the art, the method described above is based upon the well-known polymerase chain reaction (PCR) method. A skilled person would know of detailed protocols for carrying out PCR and reverse transcriptase-PCR (RT-PCR). Reviews of PCR are provided by Mullis Cold Spring Harbor Symp. Quant. Biol. 51, 263-273 (1986); Saiki et al. Bio/Technology 3, 1008-1012 (1985); and Mullis et al Meth. enzymol 155, 335-350 (1987) .

An oligonucleotide probe according to the invention has utility in detecting or quantitatively determining a nucleic acid (e.g. a DNA or RNA) encoding a human mGluR3 or mGluR4 protein according to the invention. Conventional methods for detecting or quantitatively determining a nucleic acid may be used, for example in situ hybridization, Southern blotting or Northern blotting. Accordingly, there is provided a method of detecting or quantitatively determining in a sample a target nucleic acid encoding human mGluR3 or mGluR4 according to the invention, which method comprises

(a) contacting the probe with the sample; and

(b) detecting or quantitatively determining hybridization of the probe with any target nucleic acid present in the sample.

The sample containing the target nucleotide sequence may, for example, be a tissue specimen, a tissue extract or cell extract from a patient suffering from a disease associated with abnormal human mGluR3 or mGluR4 protein activity such as a disorder of the central nervous system. Alternatively, the sample may, for example, be a sample produced as a result of a recombinant DNA procedure, in which case the sample may be recombinant cells. The target nucleic acid sequence may be a

complete human mGluR3 or mGluR4 protein coding sequence or a partial human mGluR3 or mGluR4 protein coding sequence.

A preferred method of detecting or quantitatively determining a target nucleic acid sequence in a sample comprises

(i) subjecting the sample to gel electrophoresis to separate the nucleic acids;
(ii) transferring the separated nucleic acids onto a solid support (e.g. a nitrocellulose support) by blotting; and
(iii) hybridising a probe according to the invention to the target nucleic acid sequence.

A probe can be used in an in situ hybridization procedure to locate a nucleic acid sequence encoding a human mGluR3 or mGluR4 protein of the invention. This can be done to determine the spatial distribution of human mGluR3 or mGluR4 protein coding mRNA sequences in a cell or tissue. For mRNA detection, the tissue is gently fixed so that its RNA is retained in an exposed form and the tissue is then incubated with a labelled complementary probe.

The invention includes an antibody specific for human mGluR3 or mGluR4 protein according to the invention. The antibody has utility in detecting and quantitatively determining human mGluR3 or mGluR4 protein, and hence is useful in diagnosis of diseases associated with human mGluR3 or mGluR4 protein, such as the diseases listed hereinabove. The antibody also has utility in production of human mGluR3 or mGluR4 protein by recombinant DNA procedures, for example in detection of positive clones containing a target sequence.

The antibody is preferably monoclonal, but may also be polyclonal. The antibody may be labelled. Examples of suitable antibody labels include radiolabels, biotin (which may be detected by avidin or streptavidin conjugated to

peroxidase), alkaline phosphatase and fluorescent labels (e.g. fluorescein and rhodamine). The term "antibody" is used herein to include both complete antibody molecules and fragments thereof. Preferred fragments contain at least one antigen binding site, such as Fab and F(ab')2 fragments. Humanised antibodies and fragments thereof are also included within the term "antibody".

The antibody is produced by raising antibody in a host animal against a human mGluR3 or mGluR4 protein according to the invention or an antigenic epitope (e.g. a peptide) thereof (hereinafter "the immunogen"). Methods of producing monoclonal and polyclonal antibodies are well-known. A method for producing a polyclonal antibody comprises immunising a suitable host animal, for example an experimental animal, with the immunogen and isolating immunoglobulins from the serum. The animal may therefore be inoculated with the immunogen, blood subsequently removed from the animal and the IgG fraction purified. A method for producing a monoclonal antibody comprises immortalising cells which produce the desired antibody. Hybridoma cells may be produced by fusing spleen cells from an inoculated experimental animal with tumour cells (Kohler and Milstein, Nature 256, 495-497, (1975)). The antibody may also be produced by recombinant DNA technology, for example as described in Skerra et al (1988) Science 240, 1038-1041.

An immortalized cell producing the desired antibody may be selected by a conventional procedure. The hybridomas may be grown in culture or injected intraperitoneally for formation of ascites fluid or into the blood stream of an allogenic host or immunocompromised host. Human antibody may be prepared by in vitro immunisation of human lymphocytes, followed by transformation of the lymphocytes with Epstein-Barr virus.

For the production of both moncolonal and polyclonal antibodies, the experimental animal is suitably a goat, rabbit,

rat or mouse. If desired, the immunogen may be administered as a conjugate in which the immunogen is coupled, for example via a side chain of one of the amino acid residues, to a suitably carrier. The carrier molecule is typically a physiologically acceptable carrier. The antibody obtained may be isolated and, if desired, purified.

The invention provides a method of detecting or quantitatively determining in a sample a human mGluR3 or mGluR4 protein of the invention, which method comprises

(a) contacting the sample with an antibody of the invention; and
(b) detecting or quantitatively determining the binding of the antibody.

A preferred method for detecting or quantitatively determining a human mGluR3 or mGluR4 protein is Western blotting. Such a method can comprise the steps of

(i) subjecting a sample containing a target human mGluR3 or mGluR4 protein to gel electrophoresis to separate the proteins in the sample;
(ii) transferring the separated proteins onto a solid support (e.g. a nitrocellulose support) by blotting; and
(iii) allowing an antibody according to the invention which has been labelled to bind to the target human mGluR3 or mGluR4 protein.

Preferred methods of quantitative determination are ELISA (enzyme-linked immunoassay) methods such as non-competitive ELISA methods. Typically, an ELISA method comprises the steps of

(i) immobilising on a solid support an unlabelled antibody according to the invention;

(ii) adding a sample containing the target human mGluR3 or mGluR4 protein such that the human mGluR3 or mGluR4 protein is captured by the unlabelled antibody;

(iii) adding an antibody according to the invention which has been labelled; and

(iv) quantitatively determining the amount of bound labelled antibody.

An antibody of the invention may be employed histologically for in situ detection of a human mGluR3 or mGluR4 protein, e.g. by immunofluorescence or immunoelectron micropsy. In situ detection may be accomplished by removing a histological specimen from a patient, and allowing a labelled antibody to bind to the specimen. Through use of such a procedure, it is possible to determine not only the presence of a human mGluR3 or mGluR4 protein but also its distribution.

An antibody of the invention may be used to purify a target human mGluR3 or mGluR4 protein. Conventional methods of purifying an antigen using an antibody may be used. Such methods include immunoprecipitation and immunoaffinity column methods. In an immunoaffinity column method, an antibody in accordance with the invention is coupled to the inert matrix of the column and a sample containing the target human mGluR3 or mGluR4 protein is passed down the column, such that the target human mGluR3 or mGluR4 protein is retained. The human mGluR3 or mGluR4 protein is then eluted.

The sample containing the target mGluR3 or mGluR4 protein used in the detection, determination and purification methods may be a tissue specimen, a tissue extract or a cell extract from a patient suffering from a disease associated with human mGluR3 or mGluR4 protein such as a disease listed hereinabove. Alternatively, the sample may be one produced as a result of recombinant DNA procedures, e.g. an extract of host cells.

The invention provides a pharmaceutical composition containing a pharmaceutically acceptable carrier or diluent and, as active ingredient, an antibody or polypeptide fragment of the invention.

A human mGluR3 or mGluR4 protein of the invention is useful for screening for drugs which modulate (e.g. inhibit or stimulate) the receptor. The invention includes a method for identifying a compound which binds to or modulates the activity of the human mGluR3 or mGluR4 protein, which method comprises one or more of the following steps:

(a) incubating whole cells and/or membranes prepared from cells containing the human mGluR3 or mGluR4 protein with the compound;
(b) measuring the binding of the compound to human mGluR3 or mGluR4 protein;
(c) comparing the binding with that of known ligands;
(d) measuring the effect of the compound on basal/stimulated (e.g. forskolin stimulated) adenylate cyclase activity ;
(e) comparing the ability of the compound to inhibit/stimulate human IP3 production;
(f) comparing the effect of the compound using other functional models of mGluR3 or mGluR4 activity as appropriate (e.g. microphysiometry);
(g) measuring G protein coupling by other (standard) means (e.g. binding/functional studies).

The invention also extends to agents (compounds) identified by use of a particular screen. Preferably, the agent is a chemical molecule of relatively low molecular weight, for example less than about 1000D.

Methods for identifying compounds which bind to human mGluR3 or human mGluR4 receptor generally comprise the following steps:

(a) contacting a compound with the receptor, and
(b) measuring any binding of the compound to the receptor.

Such methods for identifying compounds which bind to receptors are well-known in the art. A competitive binding assay may, for example, be used. A competitive binding assay may comprise

(a) contacting a test compound and a labelled ligand for the receptor (e.g. [^{3}H] glutamate) with the receptor,
(b) measuring the amount of the labelled ligand bound to the receptor, and
(c) comparing the amount of the labelled ligand bound to the receptor measured in step (b) with the amount of the labelled ligand that binds to the receptor in the absence of the test compound.

If less labelled ligand binds to the receptor in the presence of the test compound than in the absence of the test compound, this indicates that the test compound is competing for binding to the receptor with the ligand. The amount of the ligand bound to the receptor in step (b) may be measured indirectly; the amount of bound ligand may be determined by extrapolation from the amount of unbound ligand.

Methods for identifying compounds which modulate the activity of human mGluR3 or human mGluR4 receptor generally comprise

(a) contacting a compound with the receptor, and
(b) measuring any change in second messenger receptor activity.

The second messenger activity measured in step (b) may be basal or stimulated adenylate cyclase activity. Adenylate cyclase activity may be measured by measuring the level of cAMP, which is the product of adenylate cyclase. Typically, adenylate cyclase is stimulated using forskolin and any decrease in

adenylate cyclase activity caused or prevented by the presence of the test compound is measured. A suitable adenylate cyclase assay is described in Tanabe et al (1992) Neuron 8, 169-179.

The invention includes a combination (or complex) of human mGluR3 or human mGluR4 receptor and an agent (compound) which modulates the second messenger activity of the receptor. Examples of such agents are L-glutamate (mGluR3 and mGluR4); (2S,1'R,2'R,3'R)2-(2,3-dicarboxycyclopropyl)glycine, DCG-IV (mGluR3); L-2-amino-4-phosphonobutyrate, L-AP4 (mGluR4); trans-1-aminocyclopentane-1,3-dicarboxylate, t-ACPD (mGluR4); and quisqualate (mGluR4).

The invention also includes a method of modulating the second messenger activity of human mGluR3 or human mGluR4 receptor, which method comprises contacting the receptor with a compound which modulates the second messenger activity of the receptor. Examples of suitable compounds which modulate receptor activity are given above. The method may be carried out either on the human/animal body or outside the human or animal body (e.g. in vitro).

Brief Description of the Drawings

Figure 1 shows a comparison between human and rat mGluR3 protein sequences. The human amino acid sequence (SEQ ID NO:2) is given with differences in the rat sequence shown underneath. The seven putative transmembrane domains (TMD-I to VII) are underlined.

Figure 2 shows a comparison between human and rat mGluR4 protein sequences. The human amino acid sequence (SEQ ID NO:4) is given with differences in the rat sequence shown underneath. The seven putative transmembrane domains (TMD-I to VII) are underlined.

Figure 3 shows the effect of mGluR4 ligands on the accumulation

of forskolin-stimulated cyclic AMP. Dose response curves are given for L-amino-4-phosphonobutyrate (L-AP4), glutamate, trans-1-aminocyclopentane-1,3-dicarboxylate (t-ACPD) and quisqualate.

The following Examples illustrate the invention.

EXAMPLE 1: CLONING OF mGluR3 DNA

Construction of a Human cDNA Library

Human brain material from the median frontal cortex and amygdala from one individual was used as the source of mRNA using the Fast Track mRNA Isolation Kit (Invitrogen). The mRNA was used to construct a library containing a mixture of oligo-dT and random hexamer-primed cDNA, using the Timesaver cDNA Synthesis Kit (Pharmacia) and λZAP II arms (Stratagene).

Source of Probes

Specific regions of various mGluRs were amplified by the polymerase chain reaction (PCR) using either human brain cDNA prepared as above or in Example 2 or human genomic DNA (Stratagene) and the following consensus primers designed from the published sequences of rat mGluR1 to 5

[ACAACGACACACCCGTGGTCAA (BN29, SEQ ID NO:5)
TCCGGTCGGGAGCTCTGCTA (BN30, SEQ ID NO:6)
GCTACTCTGCCCTGCTGACCAAGAC (BN31, SEQ ID NO:7)
GCAATGCGGTTGGTCTTGGT (BN32, SEQ ID NO:8)
AAGTTCATCGGCTTCACCATGTACAC (BN33, SEQ ID NO:9)
GATGCAGGTGGTGTACATGGTGAA (BN34, SEQ ID NO:10) and
TCTCCGGTTGGAAGAGGATGATGT (BN35, SEQ ID NO:11)].

A computer-generated alignment of rat sequences of mGluR cDNAs 1 to 5 revealed a number of regions of high homology. Seven regions were used to design oligonucleotide primers such that (a) each primer contained a consensus sequence for all five

genes, (b) the highest homology of each primer was towards its 3' end and (c) the most 3' base was never a mismatch. Four primers (BN29, 30, 31 and 33) were from the coding strand and three primers (BN32, 34 and 35) were from the other strand. All seven primers came from the region of each gene which encodes the seven transmembrane domains.

PCR amplifications were performed with all nine combinations of the primers using either human genomic or human brain cDNA. Analysis of these PCR products showed that with several primer pairs, more than one mGluR subtype was amplified. Between the primer pairs all five human genes homologous to the five rat mGluR protein genes used to generate the primers were represented (Table 1). There was generally a strong bias in favour of the human mGluR3 gene. Since mGluR3 was also over-represented in PCRs involving genomic DNA, it may reflect a bias in the primers rather than the relative abundance of mGluR3 mRNA in the brain samples. When the annealing temperature in the PCR conditions was lowered to 38°C the distribution of mGluRs among the clones analysed was more even.

Unless otherwise stated the PCR conditions were: 35 cycles of 96°C for 35 sec, 56°C for 2 min and 72°C for 2 min. The PCR products were separated by electrophoresis, the bands excised and the isolated fragments cloned into pT7Blue T-vector (Novagen). The products were identified by sequencing of the cloned fragments.

Isolation of Clones

The cDNA library was screened using a mixed probe derived from inserts from cloned PCR products specific to each of human mGluR1 to 5. Hybridization was at 65°C in 10% dextran sulphate, 1% SDS, 0.1% sodium pyrophosphate, 1M Tris-Cl (pH 7.5) and 100μg/ml single stranded salmon testis DNA. The filters were washed in 1 x SSC and 1% SDS at 65oC followed by 0.1 x SSC at room temperature. Three phage clones were

isolated and stored in SM buffer (Sambrook et al, ibid). The largest of these extended from the 3' end to within 500 bp of the initiating ATG codon.

Alternative Method of Isolating Clones

One of the plates, the filter of which did not yield any of the three clones, was re-examined since its autoradiograph had a very high background. The phage from the entire plate were taken up in SM buffer and stored at 4°C in the presence of chloroform. An aliquot was used as a target for PCR using one of the pairs of consensus primers (BN30 and BN34). Electrophoresis showed that a product of the expected size (510bp) had been amplified. Phage taken from the other plates after the three mGluR3 positive phage clones had been removed did not generate this PCR product.

The PCR on the amplified phage was repeated but with samples of increasing dilution until the signal was lost. From this titration, it could be calculated that the PCR signal was derived from approximately 1 per 10^5 phage, consistent with one positive phage on the original plate of approximately 80,000 independent clones.

The positive clone was recovered from the amplified mixture using PCR after each of several rounds of dilution and replication as described by Israel , D.I., Nucl. Acids Res., 21, 2627-2631, (1993). This fourth clone was found to contain the complete coding sequence of human mGluR3. The sequence was determined for both strands of the entire insert and is shown in SEQ ID NO:1.

Sequencing of cDNA Clone

Plasmids containing the inserts were excised using helper phage R408 (Stratagene). Some sequencing was performed on such plasmid clones using flanking or internal primers. Most was on

sub-clones generated by limited exonuclease III digestion using the Nested Deletion Kit (Pharmacia).

	Primers	mGluR1	mGluR2	mGluR3	mGluR4	mGluR5
cDNA	BN29/32 (200bp)	0	0	3	0	0
	BN29/35[a] (680bp)	0	0	2	0	0
	BN30/32 (170bp)	0	1	4	0	0
	BN30/35[a] (660bp)	0	0	3	0	0
	BN31/35[a] (520bp)	0	0	4	0	0
	BN33/35[a] (180bp)	0	1	3	1	0
genomic DNA	BN29/32 (200bp)	0	1	2	0	0
	BN29/34[a] (540bp)	0	0	5	0	0
	BN30/34[a] (510bp)	1	13	1	0	0
	BN30/34[b] (510bp)	8	6	4	0	1
	BN31/34[b] (380bp)	0	10	28	1	2

Table 1. Distribution of human mGluRs amplified by different consensus primer pairs. The predicted size of each PCR product is given in brackets.
a annealing at 50°C
b annealing at 38°C

EXAMPLE 2: CLONING OF mGluR4 DNA

Construction of a Human cDNA Library

A human cDNA library was constructed in the same way as Example 1, except that human brain material from the cerebellum was used as the source of mRNA.

Source of Probes

The probes generated in Example 1 were used.

Isolation of Clones

The cDNA library was screened using an mGluR4 probe derived from inserts from cloned PCR products specific to human mGluR4. Hybridization was at 65°C in 10% dextran sulphate, 1% SDS, 0.1% sodium pyrophosphate, 1M Tris-Cl (pH 7.5) and 100 µg/ml single stranded salmon testis DNA. The filters were washed in 1 x SSC and 1% SDS at 65°C followed by 0.1 x SSC at room temperature. Phage clones were stored in SM buffer (Sambrook et al, ibid). Several clones were isolated, all at the 3' end, of which the largest extended to within 900bp of the start of the coding sequence (clone 17).

Alternative Method of Isolating Clones

New PCR primers (AP43 and AP44) (see herebelow) were designed from the published rat mGlu4 sequence. These were located in order to generate a product immediately downstream of the start of the coding sequence.

Another clone was recovered from the same cerebellum cDNA library using PCR using the new primers (AP43 and AP44 see herebelow) after each of several rounds of dilution and replication as described by Israel, D.I., Nucl. Acids Res., **21**, 2627-2631, (1993). This second clone (clone 31) was found to contain the 5' end of the coding sequence of human mGluR4, but did not overlap the 3' clone (clone 17).

5' - CTCTGCCTACTCCTCAGCCTTTA - 3' AP43, SEQ ID NO:12

5' - GCACAAAGGTCAGTGACTGCTC - 3' AP44, SEQ ID NO:13

Another pair of primers (CA396 and 400) were designed using the 3' end of clone 31 and the rat sequence in the gap between

clones 31 and 17.

AGCTCGGTCTCCATCAT CA396, SEQ ID NO:14
TGATGTCATCCTCGTTGGC CA400, SEQ ID NO:15

These primers were used to screen the same cerebellum library, again using the method of Israel. This generated clone 32 which overlapped both clones 31 and 17. Clones 17 and 32 were completely sequenced on both strands. Part of clone 31 was also sequenced on both strands to generate the coding sequence and some of the 5' untranslated sequence.

Sequencing of cDNA Clone

Sequencing was carried out in the same way as Example 1.

EXAMPLE 3: EXPRESSION OF HUMAN mGluR3

The sequence GTACAG (SEQ ID NO:1, positions 236-241) upstream of both alternative initiating ATG codons was mutated to CTCGAG (XhoI recognition sequence) by PCR. The sequence TGTGAA (SEQ ID NO:1, positions 2894-2899) downstream of the termination codon was mutated to TCTAGA (XbaI recognition sequence) by PCR. The primers AP47 and AP48 were designed for this purpose.

5' - AGAGCTCGAGAAACAGGATTCATGAAGATG - 3'AP47 (SEQ ID NO:16)
5' - GCAATCTAGAATCACAGAGATGAGGTG - 3'AP48 (SEQ ID NO:17)

The entire coding sequence of human mGluR3 was cloned between the XhoI and XbaI sites of the expression vector pSVL (Pharmacia) in which transcription of mGluR3 is under the control of the SV40 late promoter. The resulting vector is pSVLHMGR3.

DHfr$^-$ CHO cells were grown at 37°C in an atmosphere containing 5% CO_2 in glutamate-free Dulbecco's modified Eagles medium supplemented with 5% dialysed foetal calf serum, 2mM glutamine,

1mM proline, 100μM hypoxanthine and 16μM thymidine. Petri dishes (10cm) were seeded with 10^6 cells and incubated for 24 hours. The cells were washed with serum-free medium and replaced by fresh serum-free medium containing 10μg pSVLHMGR3 and 1μg pSV2 dhfr [Subramani, S. et al. Mol. Cell. Biol. Vol 1, 854 (1981)] in the presence of Transfectam (Promega). After 5 hours the medium was replaced by complete medium. After 48 hours the cells were detached from the plates with trypsin and seeded in flasks in medium without hypoxanthine/thymidine. After 1-2 weeks surviving colonies which were dhfr$^+$ were pooled and clonal colonies were obtained by dilution cloning.

Several clones were selected and transferred to individual flasks. Each culture was assayed for mGluR3 receptor activity by measuring the inhibition of forskolin-stimulated cyclic AMP in response to the agonist trans-1-aminocyclopentane-1,3-dicarboxylate (t-ACPD).

Forskolin-stimulated cyclic AMP assay: Approximately 2 x 10^5 cells were seeded in each well of a 12 well microtitre plate. After incubation at 37°C for 24 hours the cells were exposed to 100μM 3-isobutyl-1-methylxanthine (phosphodiesterase inhibitor) at 37°C for 20 minutes followed by both the ligand and 10μM forskolin for 10 minutes. The medium was removed and ice-cold ethanol was added to the cells. After 2 hours at room temperature the supernatant was transferred to fresh tubes and the ethanol was removed by evaporation. The accumulation of cyclic AMP was measured in each tube using a scintillation proximity assay (Amersham).

EXAMPLE 4: EXPRESSION OF HUMAN mGluR4

The complete mGluR4 coding sequence was constructed from clones 17, 31 and 32, utilising the restriction endonucleases EcoRI (SEQ ID NO:3, position 290-295), FspI (SEQ ID NO:3, position 1615-1620) and NcoI (SEQ ID NO:3, position 2915-2920). A unique BamHI site was introduced downstream of both the TAG

stop codon and of the NcoI site by subcloning the EcoRI-FspI fragment from clone 32 and the FspI-NcoI fragment from clone 17 into pSL1190 (Pharmacia). The sequence TCTAGG in clone 31 (SEQ ID NO:3, positions 155-160) upstream of the initiating ATG codon was mutated to TCTAGA (XbaI recognition sequence) by PCR using primers CA587 and CA565.

5' - GGGTGTCTAGAGATTTCCGAG - 3' CA587 (SEQ ID NO:18)
5' - AGAGGCGAAGGATGTTG - 3' CA565 (SEQ ID NO:19)

The entire coding sequence was cloned as an XbaI-BamHI fragment into the expression vector pSVL (Pharmacia) in which transcription of mGluR4 is under the control of the SV40 late promoter. The resulting vector is pSVLHMGR4.

Dhrf$^-$ CHO cells were grown at 37°C in an atmosphere containing 5% CO_2 in glutamate-free Dulbecco's modified Eagles medium supplemented with 5% dialysed foetal calf serum, 2mM glutamine, 1mM proline, 100μM hypoxanthine and 16μM thymidine. Petri dishes (10cm) were seeded with 10^6 cells and incubated for 24 hours. The cells were washed with serum-free medium and replaced by fresh serum-free medium containing 10μg pSVLHMGR4 plus pSV2 dhfr [Subramani S. et al Mol. Cell. Biol. Vol 1, 854 (1981)] in a range between 10ng and 1μg in the presence of Transfectam (Promega). After 5 hours the medium was replaced by complete medium. After 48 hours the cells were detached from the plates with trypsin and seeded in 15cm petri dishes in medium without hypoxanthine/thymidine. After 1-2 weeks surviving colonies which were dhfr$^+$ were selected and transferred to individual flasks. Each selected culture was assayed for mGluR4 receptor activity by measuring the inhibition of forskolin-stimulated cyclic AMP in response to the agonist L-2-amino-4-phosphonobutyrate (L-AP4). Cyclic AMP assays were performed as for mGluR3. Expressing colonies were then dilution cloned to obtain clonal cell lines.

EXAMPLE 5: INTERACTION BETWEEN HUMAN mGluR4 AND LIGANDS

The effect of several ligands on human mGluR4 was investigated by measuring the accumulation of forskolin-stimulated cyclic AMP. Dose response curves are given in Figure 3 for L-2-amino-4-phosphonobutyrate (L-AP4), glutamate, trans-1-aminocyclopentane-1,3-dicarboxylate (t-ACPD) and quisqualate.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: The Wellcome Foundation Limited
(B) STREET: Unicorn House, 160 Euston Road
(C) CITY: London
(D) STATE: not applicable
(E) COUNTRY: United Kingdom
(F) POSTAL CODE (ZIP): NW1 2BP

(A) NAME: MAKOFF, Andrew Joseph
(B) STREET: Langley Court
(C) CITY: Beckenham
(D) STATE: Kent
(E) COUNTRY: United Kingdom
(F) POSTAL CODE (ZIP): BR3 3BS

(ii) TITLE OF INVENTION: HUMAN GLUTAMATE RECEPTOR PROTEINS

(iii) NUMBER OF SEQUENCES: 19

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3410 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:259..2889

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CTTTTGTGTC GGATGAGGAG GACCAACCAT GAGCCAGAGC CCGGGTGCAG GCTCACCGCC 60

GCCGCTGCCA CCGCGGTCAG CTCCAGTTCC TGCCAGGAGT TGTCGGTGCG AGGAATTTTG 120

TGACAGGCTC TGTTAGTCTG TTCCTCCCTT ATTTGAAGGA CAGGCCAAAG ATCCAGTTTG 180

GAAATGAGAG AGGACTAGCA TGACACATTG GCTCCACCAT TGATATCTCC CAGAGGTACA 240

```
GAAACAGGAT TCATGAAG ATG TTG ACA AGA CTG CAA GTT CTT ACC TTA GCT          291
                    Met Leu Thr Arg Leu Gln Val Leu Thr Leu Ala
                      1               5                  10

TTG TTT TCA AAG GGA TTT TTA CTC TCT TTA GGG GAC CAT AAC TTT CTA          339
Leu Phe Ser Lys Gly Phe Leu Leu Ser Leu Gly Asp His Asn Phe Leu
            15                  20                  25

AGG AGA GAG ATT AAA ATA GAA GGT GAC CTT GTT TTA GGG GGC CTG TTT          387
Arg Arg Glu Ile Lys Ile Glu Gly Asp Leu Val Leu Gly Gly Leu Phe
        30                  35                  40

CCT ATT AAC GAA AAA GGC ACT GGA ACT GAA GAA TGT GGG CGA ATC AAT          435
Pro Ile Asn Glu Lys Gly Thr Gly Thr Glu Glu Cys Gly Arg Ile Asn
    45                  50                  55

GAA GAC CGA GGG ATT CAA CGC CTG GAA GCC ATG TTG TTT GCT ATT GAT          483
Glu Asp Arg Gly Ile Gln Arg Leu Glu Ala Met Leu Phe Ala Ile Asp
60                  65                  70                  75

GAA ATC AAC AAA GAT GAT TAC TTG CTA CCA GGA GTG AAG TTG GGT GTT          531
Glu Ile Asn Lys Asp Asp Tyr Leu Leu Pro Gly Val Lys Leu Gly Val
                80                  85                  90

CAC ATT TTG GAT ACA TGT TCA AGG GAT ACC TAT GCA TTG GAG CAA TCA          579
His Ile Leu Asp Thr Cys Ser Arg Asp Thr Tyr Ala Leu Glu Gln Ser
            95                  100                 105

CTG GAG TTT GTC AGG GCA TCT TTG ACA AAA GTG GAT GAA GCT GAG TAT          627
Leu Glu Phe Val Arg Ala Ser Leu Thr Lys Val Asp Glu Ala Glu Tyr
        110                 115                 120

ATG TGT CCT GAT GGA TCC TAT GCC ATT CAA GAA AAC ATC CCA CTT CTC          675
Met Cys Pro Asp Gly Ser Tyr Ala Ile Gln Glu Asn Ile Pro Leu Leu
    125                 130                 135

ATT GCA GGG GTC ATT GGT GGC TCT TAT AGC AGT GTT TCC ATA CAG GTG          723
Ile Ala Gly Val Ile Gly Gly Ser Tyr Ser Ser Val Ser Ile Gln Val
140                 145                 150                 155

GCA AAC CTG CTG CGG CTC TTC CAG ATC CCT CAG ATC AGC TAC GCA TCC          771
Ala Asn Leu Leu Arg Leu Phe Gln Ile Pro Gln Ile Ser Tyr Ala Ser
                160                 165                 170

ACC AGC GCC AAA CTC AGT GAT AAG TCG CGC TAT GAT TAC TTT GCC AGG          819
Thr Ser Ala Lys Leu Ser Asp Lys Ser Arg Tyr Asp Tyr Phe Ala Arg
            175                 180                 185

ACC GTG CCC CCC GAC TTC TAC CAG GCC AAA GCC ATG GCT GAG ATC TTG          867
Thr Val Pro Pro Asp Phe Tyr Gln Ala Lys Ala Met Ala Glu Ile Leu
        190                 195                 200

CGC TTC TTC AAC TGG ACC TAC GTG TCC ACA GTA GCC TCC GAG GGT GAT          915
Arg Phe Phe Asn Trp Thr Tyr Val Ser Thr Val Ala Ser Glu Gly Asp
    205                 210                 215
```

```
TAC GGG GAG ACA GGG ATC GAG GCC TTC GAG CAG GAA GCC CGC CTG CGC      963
Tyr Gly Glu Thr Gly Ile Glu Ala Phe Glu Gln Glu Ala Arg Leu Arg
220                 225                 230                 235

AAC ATC TGC ATC GCT ACG GCG GAG AAG GTG GGC CGC TCC AAC ATC CGC     1011
Asn Ile Cys Ile Ala Thr Ala Glu Lys Val Gly Arg Ser Asn Ile Arg
                240                 245                 250

AAG TCC TAC GAC AGC GTG ATC CGA GAA CTG TTG CAG AAG CCC AAC GCG     1059
Lys Ser Tyr Asp Ser Val Ile Arg Glu Leu Leu Gln Lys Pro Asn Ala
            255                 260                 265

CGC GTC GTG GTC CTC TTC ATG CGC AGC GAC GAC TCG CGG GAG CTC ATT     1107
Arg Val Val Val Leu Phe Met Arg Ser Asp Asp Ser Arg Glu Leu Ile
        270                 275                 280

GCA GCC GCC AGC CGC GCC AAT GCC TCC TTC ACC TGG GTG GCC AGC GAC     1155
Ala Ala Ala Ser Arg Ala Asn Ala Ser Phe Thr Trp Val Ala Ser Asp
    285                 290                 295

GGC TGG GGC GCG CAG GAG AGC ATC ATC AAG GGC AGC GAG CAT GTG GCC     1203
Gly Trp Gly Ala Gln Glu Ser Ile Ile Lys Gly Ser Glu His Val Ala
300                 305                 310                 315

TAC GGC GCC ATC ACC CTG GAG CTG GCC TCC CAG CCT GTC CGC CAG TTC     1251
Tyr Gly Ala Ile Thr Leu Glu Leu Ala Ser Gln Pro Val Arg Gln Phe
                320                 325                 330

GAC CGC TAC TTC CAG AGC CTC AAC CCC TAC AAC AAC CAC CGC AAC CCC     1299
Asp Arg Tyr Phe Gln Ser Leu Asn Pro Tyr Asn Asn His Arg Asn Pro
            335                 340                 345

TGG TTC CGG GAC TTC TGG GAG CAA AAG TTT CAG TGC AGC CTC CAG AAC     1347
Trp Phe Arg Asp Phe Trp Glu Gln Lys Phe Gln Cys Ser Leu Gln Asn
        350                 355                 360

AAA CGC AAC CAC AGG CGC GTC TGC GAC AAG CAC CTG GCC ATC GAC AGC     1395
Lys Arg Asn His Arg Arg Val Cys Asp Lys His Leu Ala Ile Asp Ser
    365                 370                 375

AGC AAC TAC GAG CAA GAG TCC AAG ATC ATG TTT GTG GTG AAC GCG GTG     1443
Ser Asn Tyr Glu Gln Glu Ser Lys Ile Met Phe Val Val Asn Ala Val
380                 385                 390                 395

TAT GCC ATG GCC CAC GCT TTG CAC AAA ATG CAG CGC ACC CTC TGT CCC     1491
Tyr Ala Met Ala His Ala Leu His Lys Met Gln Arg Thr Leu Cys Pro
                400                 405                 410

AAC ACT ACC AAG CTT TGT GAT GCT ATG AAG ATC CTG GAT GGG AAG AAG     1539
Asn Thr Thr Lys Leu Cys Asp Ala Met Lys Ile Leu Asp Gly Lys Lys
            415                 420                 425

TTG TAC AAG GAT TAC TTG CTG AAA ATC AAC TTC ACG GCT CCA TTC AAC     1587
Leu Tyr Lys Asp Tyr Leu Leu Lys Ile Asn Phe Thr Ala Pro Phe Asn
        430                 435                 440
```

```
CCA AAT AAA GAT GCA GAT AGC ATA GTC AAG TTT GAC ACT TTT GGA GAT      1635
Pro Asn Lys Asp Ala Asp Ser Ile Val Lys Phe Asp Thr Phe Gly Asp
    445                 450                 455

GGA ATG GGG CGA TAC AAC GTG TTC AAT TTC CAA AAT GTA GGT GGA AAG      1683
Gly Met Gly Arg Tyr Asn Val Phe Asn Phe Gln Asn Val Gly Gly Lys
460                 465                 470                 475

TAT TCC TAC TTG AAA GTT GGT CAC TGG GCA GAA ACC TTA TCG CTA GAT      1731
Tyr Ser Tyr Leu Lys Val Gly His Trp Ala Glu Thr Leu Ser Leu Asp
                480                 485                 490

GTC AAC TCT ATC CAC TGG TCC CGG AAC TCA GTC CCC ACT TCC CAG TGC      1779
Val Asn Ser Ile His Trp Ser Arg Asn Ser Val Pro Thr Ser Gln Cys
            495                 500                 505

AGC GAC CCC TGT GCC CCC AAT GAA ATG AAG AAT ATG CAA CCA GGG GAT      1827
Ser Asp Pro Cys Ala Pro Asn Glu Met Lys Asn Met Gln Pro Gly Asp
        510                 515                 520

GTC TGC TGC TGG ATT TGC ATC CCC TGT GAA CCC TAC GAA TAC CTG GCT      1875
Val Cys Cys Trp Ile Cys Ile Pro Cys Glu Pro Tyr Glu Tyr Leu Ala
    525                 530                 535

GAT GAG TTT ACC TGT ATG GAT TGT GGG TCT GGA CAG TGG CCC ACT GCA      1923
Asp Glu Phe Thr Cys Met Asp Cys Gly Ser Gly Gln Trp Pro Thr Ala
540                 545                 550                 555

GAC CTA ACT GGA TGC TAT GAC CTT CCT GAG GAC TAC ATC AGG TGG GAA      1971
Asp Leu Thr Gly Cys Tyr Asp Leu Pro Glu Asp Tyr Ile Arg Trp Glu
                560                 565                 570

GAC GCC TGG GCC ATT GGC CCA GTC ACC ATT GCC TGT CTG GGT TTT ATG      2019
Asp Ala Trp Ala Ile Gly Pro Val Thr Ile Ala Cys Leu Gly Phe Met
            575                 580                 585

TGT ACA TGC ATG GTT GTA ACT GTT TTT ATC AAG CAC AAC AAC ACA CCC      2067
Cys Thr Cys Met Val Val Thr Val Phe Ile Lys His Asn Asn Thr Pro
        590                 595                 600

TTG GTC AAA GCA TCG GGC CGA GAA CTC TGC TAC ATC TTA TTG TTT GGG      2115
Leu Val Lys Ala Ser Gly Arg Glu Leu Cys Tyr Ile Leu Leu Phe Gly
    605                 610                 615

GTT GGC CTG TCA TAC TGC ATG ACA TTC TTC TTC ATT GCC AAG CCA TCA      2163
Val Gly Leu Ser Tyr Cys Met Thr Phe Phe Phe Ile Ala Lys Pro Ser
620                 625                 630                 635

CCA GTC ATC TGT GCA TTG CGC CGA CTC GGG CTG GGG AGT TCC TTC GCT      2211
Pro Val Ile Cys Ala Leu Arg Arg Leu Gly Leu Gly Ser Ser Phe Ala
                640                 645                 650

ATC TGT TAC TCA GCC CTG CTG ACC AAG ACA AAC TGC ATT GCC CGC ATC      2259
Ile Cys Tyr Ser Ala Leu Leu Thr Lys Thr Asn Cys Ile Ala Arg Ile
            655                 660                 665
```

```
TTC GAT GGG GTC AAG AAT GGC GCT CAG AGG CCA AAA TTC ATC AGC CCC      2307
Phe Asp Gly Val Lys Asn Gly Ala Gln Arg Pro Lys Phe Ile Ser Pro
        670                 675                 680

AGT TCT CAG GTT TTC ATC TGC CTG GGT CTG ATC CTG GTG CAA ATT GTG      2355
Ser Ser Gln Val Phe Ile Cys Leu Gly Leu Ile Leu Val Gln Ile Val
    685                 690                 695

ATG GTG TCT GTG TGG CTC ATC CTG GAG GCC CCA GGC ACC AGG AGG TAT      2403
Met Val Ser Val Trp Leu Ile Leu Glu Ala Pro Gly Thr Arg Arg Tyr
700                 705                 710                 715

ACC CTT GCA GAG AAG CGG GAA ACA GTC ATC CTA AAA TGC AAT GTC AAA      2451
Thr Leu Ala Glu Lys Arg Glu Thr Val Ile Leu Lys Cys Asn Val Lys
                720                 725                 730

GAT TCC AGC ATG TTG ATC TCT CTT ACC TAC GAT GTG ATC CTG GTG ATC      2499
Asp Ser Ser Met Leu Ile Ser Leu Thr Tyr Asp Val Ile Leu Val Ile
            735                 740                 745

TTA TGC ACT GTG TAC GCC TTC AAA ACG CGG AAG TGC CCA GAA AAT TTC      2547
Leu Cys Thr Val Tyr Ala Phe Lys Thr Arg Lys Cys Pro Glu Asn Phe
        750                 755                 760

AAC GAA GCT AAG TTC ATA GGT TTT ACC ATG TAC ACC ACG TGC ATC ATC      2595
Asn Glu Ala Lys Phe Ile Gly Phe Thr Met Tyr Thr Thr Cys Ile Ile
    765                 770                 775

TGG TTG GCC TTC CTC CCT ATA TTT TAT GTG ACA TCA AGT GAC TAC AGA      2643
Trp Leu Ala Phe Leu Pro Ile Phe Tyr Val Thr Ser Ser Asp Tyr Arg
780                 785                 790                 795

GTG CAG ACG ACA ACC ATG TGC ATC TCT GTC AGC CTG AGT GGC TTT GTG      2691
Val Gln Thr Thr Thr Met Cys Ile Ser Val Ser Leu Ser Gly Phe Val
                800                 805                 810

GTC TTG GGC TGT TTG TTT GCA CCC AAG GTT CAC ATC ATC CTG TTT CAA      2739
Val Leu Gly Cys Leu Phe Ala Pro Lys Val His Ile Ile Leu Phe Gln
            815                 820                 825

CCC CAG AAG AAT GTT GTC ACA CAC AGA CTG CAC CTC AAC AGG TTC AGT      2787
Pro Gln Lys Asn Val Val Thr His Arg Leu His Leu Asn Arg Phe Ser
        830                 835                 840

GTC AGT GGA ACT GGG ACC ACA TAC TCT CAG TCC TCT GCA AGC ACG TAT      2835
Val Ser Gly Thr Gly Thr Thr Tyr Ser Gln Ser Ser Ala Ser Thr Tyr
    845                 850                 855

GTG CCA ACG GTG TGC AAT GGG CGG GAA GTC CTC GAC TCC ACC ACC TCA      2883
Val Pro Thr Val Cys Asn Gly Arg Glu Val Leu Asp Ser Thr Thr Ser
860                 865                 870                 875

TCT CTG TGATTGTGAA TTGCAGTTCA GTTCTTGTGT TTTTAGACTG TTAGACAAAA      2939
Ser Leu
```

GTGCTCACGT GCAGCTCCAG AATATGGAAA CAGAGCAAAA GAACAACCCT AGTACCTTTT 2999

TTTAGAAACA GTACGATAAA TTATTTTTGA GGACTGTATA TAGTGATGTG CTAGAACTTT 3059

CTAGGCTGAG TCTAGTGCCC CTATTATTAA CAATTCCCCC AGAACATGGA AATAACCATT 3119

GTTTACAGAG CTGAGCATTG GTGACAGGGT CTGACATGGT CAGTCTACTA AAAAACAAAA 3179

AAAAAAAACA AAAAAAAAAA AACAAAAGAA AAAAATAAAA ATACGGTGGC AATATTATGT 3239

AACCTTTTTT CCTATGAAGT TTTTTGTAGG TCCTTGTTGT AACTAATTTA GGATGAGTTT 3299

CTATGTTGTA TATTAAAGTT ACATTATGTG TAACAGATTG ATTTTCTCAG CACAAAATAA 3359

AAAGCATCTG TATTAATGTA AAGATACTGA GAATAAAACC TTCAAGGTTT T 3410

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 877 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Leu Thr Arg Leu Gln Val Leu Thr Leu Ala Leu Phe Ser Lys Gly
1 5 10 15

Phe Leu Leu Ser Leu Gly Asp His Asn Phe Leu Arg Arg Glu Ile Lys
20 25 30

Ile Glu Gly Asp Leu Val Leu Gly Gly Leu Phe Pro Ile Asn Glu Lys
35 40 45

Gly Thr Gly Thr Glu Glu Cys Gly Arg Ile Asn Glu Asp Arg Gly Ile
50 55 60

Gln Arg Leu Glu Ala Met Leu Phe Ala Ile Asp Glu Ile Asn Lys Asp
65 70 75 80

Asp Tyr Leu Leu Pro Gly Val Lys Leu Gly Val His Ile Leu Asp Thr
85 90 95

Cys Ser Arg Asp Thr Tyr Ala Leu Glu Gln Ser Leu Glu Phe Val Arg
100 105 110

Ala Ser Leu Thr Lys Val Asp Glu Ala Glu Tyr Met Cys Pro Asp Gly
115 120 125

Ser Tyr Ala Ile Gln Glu Asn Ile Pro Leu Leu Ile Ala Gly Val Ile
130 135 140

Gly Gly Ser Tyr Ser Ser Val Ser Ile Gln Val Ala Asn Leu Leu Arg
145 150 155 160

Leu Phe Gln Ile Pro Gln Ile Ser Tyr Ala Ser Thr Ser Ala Lys Leu
165 170 175

Ser Asp Lys Ser Arg Tyr Asp Tyr Phe Ala Arg Thr Val Pro Pro Asp
180 185 190

Phe Tyr Gln Ala Lys Ala Met Ala Glu Ile Leu Arg Phe Phe Asn Trp
195 200 205

Thr Tyr Val Ser Thr Val Ala Ser Glu Gly Asp Tyr Gly Glu Thr Gly
210 215 220

Ile Glu Ala Phe Glu Gln Glu Ala Arg Leu Arg Asn Ile Cys Ile Ala
225 230 235 240

Thr Ala Glu Lys Val Gly Arg Ser Asn Ile Arg Lys Ser Tyr Asp Ser
245 250 255

Val Ile Arg Glu Leu Leu Gln Lys Pro Asn Ala Arg Val Val Val Leu
260 265 270

Phe Met Arg Ser Asp Asp Ser Arg Glu Leu Ile Ala Ala Ala Ser Arg
275 280 285

Ala Asn Ala Ser Phe Thr Trp Val Ala Ser Asp Gly Trp Gly Ala Gln
290 295 300

Glu Ser Ile Ile Lys Gly Ser Glu His Val Ala Tyr Gly Ala Ile Thr
305 310 315 320

Leu Glu Leu Ala Ser Gln Pro Val Arg Gln Phe Asp Arg Tyr Phe Gln
325 330 335

Ser Leu Asn Pro Tyr Asn Asn His Arg Asn Pro Trp Phe Arg Asp Phe
340 345 350

Trp Glu Gln Lys Phe Gln Cys Ser Leu Gln Asn Lys Arg Asn His Arg
355 360 365

Arg Val Cys Asp Lys His Leu Ala Ile Asp Ser Ser Asn Tyr Glu Gln
370 375 380

Glu Ser Lys Ile Met Phe Val Val Asn Ala Val Tyr Ala Met Ala His
385 390 395 400

Ala Leu His Lys Met Gln Arg Thr Leu Cys Pro Asn Thr Thr Lys Leu
405 410 415

Cys Asp Ala Met Lys Ile Leu Asp Gly Lys Lys Leu Tyr Lys Asp Tyr
420 425 430

Leu Leu Lys Ile Asn Phe Thr Ala Pro Phe Asn Pro Asn Lys Asp Ala
435 440 445

Asp Ser Ile Val Lys Phe Asp Thr Phe Gly Asp Gly Met Gly Arg Tyr
450 455 460

Asn Val Phe Asn Phe Gln Asn Val Gly Gly Lys Tyr Ser Tyr Leu Lys
465 470 475 480

Val Gly His Trp Ala Glu Thr Leu Ser Leu Asp Val Asn Ser Ile His
485 490 495

Trp Ser Arg Asn Ser Val Pro Thr Ser Gln Cys Ser Asp Pro Cys Ala
500 505 510

Pro Asn Glu Met Lys Asn Met Gln Pro Gly Asp Val Cys Cys Trp Ile
515 520 525

Cys Ile Pro Cys Glu Pro Tyr Glu Tyr Leu Ala Asp Glu Phe Thr Cys
530 535 540

Met Asp Cys Gly Ser Gly Gln Trp Pro Thr Ala Asp Leu Thr Gly Cys
545 550 555 560

Tyr Asp Leu Pro Glu Asp Tyr Ile Arg Trp Glu Asp Ala Trp Ala Ile
565 570 575

Gly Pro Val Thr Ile Ala Cys Leu Gly Phe Met Cys Thr Cys Met Val
580 585 590

Val Thr Val Phe Ile Lys His Asn Asn Thr Pro Leu Val Lys Ala Ser
595 600 605

Gly Arg Glu Leu Cys Tyr Ile Leu Leu Phe Gly Val Gly Leu Ser Tyr
610 615 620

Cys Met Thr Phe Phe Phe Ile Ala Lys Pro Ser Pro Val Ile Cys Ala
625 630 635 640

Leu Arg Arg Leu Gly Leu Gly Ser Ser Phe Ala Ile Cys Tyr Ser Ala
645 650 655

Leu Leu Thr Lys Thr Asn Cys Ile Ala Arg Ile Phe Asp Gly Val Lys
660 665 670

Asn Gly Ala Gln Arg Pro Lys Phe Ile Ser Pro Ser Ser Gln Val Phe
675 680 685

Ile Cys Leu Gly Leu Ile Leu Val Gln Ile Val Met Val Ser Val Trp
690 695 700

Leu Ile Leu Glu Ala Pro Gly Thr Arg Arg Tyr Thr Leu Ala Glu Lys
705 710 715 720

Arg Glu Thr Val Ile Leu Lys Cys Asn Val Lys Asp Ser Ser Met Leu
725 730 735

Ile Ser Leu Thr Tyr Asp Val Ile Leu Val Ile Leu Cys Thr Val Tyr
740 745 750

Ala Phe Lys Thr Arg Lys Cys Pro Glu Asn Phe Asn Glu Ala Lys Phe
755 760 765

```
Ile Gly Phe Thr Met Tyr Thr Thr Cys Ile Ile Trp Leu Ala Phe Leu
    770                 775                 780

Pro Ile Phe Tyr Val Thr Ser Ser Asp Tyr Arg Val Gln Thr Thr Thr
785                 790                 795                 800

Met Cys Ile Ser Val Ser Leu Ser Gly Phe Val Val Leu Gly Cys Leu
                805                 810                 815

Phe Ala Pro Lys Val His Ile Ile Leu Phe Gln Pro Gln Lys Asn Val
            820                 825                 830

Val Thr His Arg Leu His Leu Asn Arg Phe Ser Val Ser Gly Thr Gly
        835                 840                 845

Thr Thr Tyr Ser Gln Ser Ser Ala Ser Thr Tyr Val Pro Thr Val Cys
    850                 855                 860

Asn Gly Arg Glu Val Leu Asp Ser Thr Thr Ser Ser Leu
865                 870                 875
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3884 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:171..2906

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
CCGAGTGACA AGGAGGTGGG AGAGGGTAGC AGCATGGGCT ACGCGGTTGG CTGCCCTCAG      60

TCCCCCTGCT GCTGAAGCTG CCCTGCCCAT GCCCACCCAG GCCGTGGGGC CAGGGGCCTG     120

CCAGGGCTAG GAGTGGGCCT GCCGTTCATG GGTCTCTAGG GATTTCCGAG ATG CCT        176
                                                          Met Pro

GGG AAG AGA GGC TTG GGC TGG TGG TGG GCC CGG CTG CCC CTT TGC CTG       224
Gly Lys Arg Gly Leu Gly Trp Trp Trp Ala Arg Leu Pro Leu Cys Leu
880                 885                 890                 895

CTC CTC AGC CTT TAC GGC CCC TGG ATG CCT TCC TCC CTG GGA AAG CCC       272
Leu Leu Ser Leu Tyr Gly Pro Trp Met Pro Ser Ser Leu Gly Lys Pro
                900                 905                 910

AAA GGC CAC CCT CAC ATG AAT TCC ATC CGC ATA GAT GGG GAC ATC ACA       320
Lys Gly His Pro His Met Asn Ser Ile Arg Ile Asp Gly Asp Ile Thr
```

```
                915                 920                 925

CTG GGA GGC CTG TTC CCG GTG CAT GGC CGG GGC TCA GAG GGC AAG CCC      368
Leu Gly Gly Leu Phe Pro Val His Gly Arg Gly Ser Glu Gly Lys Pro
        930                 935                 940

TGT GGA GAA CTT AAG AAG GAA AAG GGC ATC CAC CGG CTG GAG GCC ATG      416
Cys Gly Glu Leu Lys Lys Glu Lys Gly Ile His Arg Leu Glu Ala Met
    945                 950                 955

CTG TTC GCC CTG GAT CGC ATC AAC AAC GAC CCG GAC CTG CTG CCT AAC      464
Leu Phe Ala Leu Asp Arg Ile Asn Asn Asp Pro Asp Leu Leu Pro Asn
960                 965                 970                 975

ATC ACG CTG GGC GCC CGC ATT CTG GAC ACC TGC TCC AGG GAC ACC CAT      512
Ile Thr Leu Gly Ala Arg Ile Leu Asp Thr Cys Ser Arg Asp Thr His
                980                 985                 990

GCC CTC GAG CAG TCG CTG ACC TTT GTG CAG GCG CTC ATC GAG AAG GAT      560
Ala Leu Glu Gln Ser Leu Thr Phe Val Gln Ala Leu Ile Glu Lys Asp
            995                1000                1005

GGC ACA GAG GTC CGC TGT GGC AGT GGC GGC CCA CCC ATC ATC ACC AAG      608
Gly Thr Glu Val Arg Cys Gly Ser Gly Gly Pro Pro Ile Ile Thr Lys
        1010                1015                1020

CCT GAA CGT GTG GTG GGT GTC ATC GGT GCT TCA GGG AGC TCG GTC TCC      656
Pro Glu Arg Val Val Gly Val Ile Gly Ala Ser Gly Ser Ser Val Ser
    1025                1030                1035

ATC ATG GTG GCC AAC ATC CTT CGC CTC TTC AAG ATA CCC CAG ATC AGC      704
Ile Met Val Ala Asn Ile Leu Arg Leu Phe Lys Ile Pro Gln Ile Ser
1040                1045                1050                1055

TAC GCC TCC ACA GCG CCA GAC CTG AGT GAC AAC AGC CGC TAC GAC TTC      752
Tyr Ala Ser Thr Ala Pro Asp Leu Ser Asp Asn Ser Arg Tyr Asp Phe
                1060                1065                1070

TTC TCC CGC GTG GTG CCC TCG GAC ACG TAC CAG GCC CAG GCC ATG GTG      800
Phe Ser Arg Val Val Pro Ser Asp Thr Tyr Gln Ala Gln Ala Met Val
            1075                1080                1085

GAC ATC GTC CGT GCC CTC AAG TGG AAC TAT GTG TCC ACA GTG GCC TCG      848
Asp Ile Val Arg Ala Leu Lys Trp Asn Tyr Val Ser Thr Val Ala Ser
        1090                1095                1100

GAG GGC AGC TAT GGT GAG AGC GGT GTG GAG GCC TTC ATC CAG AAG TCC      896
Glu Gly Ser Tyr Gly Glu Ser Gly Val Glu Ala Phe Ile Gln Lys Ser
    1105                1110                1115

CGT GAG GAC GGG GGC GTG TGC ATC GCC CAG TCG GTG AAG ATA CCA CGG      944
Arg Glu Asp Gly Gly Val Cys Ile Ala Gln Ser Val Lys Ile Pro Arg
1120                1125                1130                1135

GAG CCC AAG GCA GGC GAG TTC GAC AAG ATC ATC CGC CGC CTC CTG GAG      992
Glu Pro Lys Ala Gly Glu Phe Asp Lys Ile Ile Arg Arg Leu Leu Glu
```

```
                    1140                1145                1150

ACT TCG AAC GCC AGG GCA GTC ATC ATC TTT GCC AAC GAG GAT GAC ATC      1040
Thr Ser Asn Ala Arg Ala Val Ile Ile Phe Ala Asn Glu Asp Asp Ile
            1155                1160                1165

AGG CGT GTG CTG GAG GCA GCA CGA AGG GCC AAC CAG ACA GGC CAT TTC      1088
Arg Arg Val Leu Glu Ala Ala Arg Arg Ala Asn Gln Thr Gly His Phe
        1170                1175                1180

TTC TGG ATG GGC TCT GAC AGC TGG GGC TCC AAG ATT GCA CCT GTG CTG      1136
Phe Trp Met Gly Ser Asp Ser Trp Gly Ser Lys Ile Ala Pro Val Leu
    1185                1190                1195

CAC CTG GAG GAG GTG GCT GAG GGT GCT GTC ACG ATC CTC CCC AAG AGG      1184
His Leu Glu Glu Val Ala Glu Gly Ala Val Thr Ile Leu Pro Lys Arg
1200                1205                1210                1215

ATG TCC GTA CGA GGC TTC GAC CGC TAC TTC TCC AGC CGC ACG CTG GAC      1232
Met Ser Val Arg Gly Phe Asp Arg Tyr Phe Ser Ser Arg Thr Leu Asp
                1220                1225                1230

AAC AAC CGG CGC AAC ATC TGG TTT GCC GAG TTC TGG GAG GAC AAC TTC      1280
Asn Asn Arg Arg Asn Ile Trp Phe Ala Glu Phe Trp Glu Asp Asn Phe
            1235                1240                1245

CAC TGC AAG CTG AGC CGC CAC GCC CTC AAG AAG GGC AGC CAC GTC AAG      1328
His Cys Lys Leu Ser Arg His Ala Leu Lys Lys Gly Ser His Val Lys
        1250                1255                1260

AAG TGC ACC AAC CGT GAG CGA ATT GGG CAG GAT TCA GCT TAT GAG CAG      1376
Lys Cys Thr Asn Arg Glu Arg Ile Gly Gln Asp Ser Ala Tyr Glu Gln
    1265                1270                1275

GAG GGG AAG GTG CAG TTT GTG ATC GAT GCC GTG TAC GCC ATG GGC CAC      1424
Glu Gly Lys Val Gln Phe Val Ile Asp Ala Val Tyr Ala Met Gly His
1280                1285                1290                1295

GCG CTG CAC GCC ATG CAC CGT GAC CTG TGT CCC GGC CGC GTG GGG CTC      1472
Ala Leu His Ala Met His Arg Asp Leu Cys Pro Gly Arg Val Gly Leu
                1300                1305                1310

TGC CCG CGC ATG GAC CCT GTA GAT GGC ACC CAG CTG CTT AAG TAC ATC      1520
Cys Pro Arg Met Asp Pro Val Asp Gly Thr Gln Leu Leu Lys Tyr Ile
            1315                1320                1325

CGA AAC GTC AAC TTC TCA GGC ATC GCA GGG AAC CCT GTG ACC TTC AAT      1568
Arg Asn Val Asn Phe Ser Gly Ile Ala Gly Asn Pro Val Thr Phe Asn
        1330                1335                1340

GAG AAT GGA GAT GCG CCT GGG CGC TAT GAC ATC TAC CAA TAC CAG CTG      1616
Glu Asn Gly Asp Ala Pro Gly Arg Tyr Asp Ile Tyr Gln Tyr Gln Leu
    1345                1350                1355

CGC AAC GAT TCT GCC GAG TAC AAG GTC ATT GGC TCC TGG ACT GAC CAC      1664
Arg Asn Asp Ser Ala Glu Tyr Lys Val Ile Gly Ser Trp Thr Asp His
```

```
1360                1365                1370                1375

CTG CAC CTT AGA ATA GAG CGG ATG CAC TGG CCG GGG AGC GGG CAG CAG      1712
Leu His Leu Arg Ile Glu Arg Met His Trp Pro Gly Ser Gly Gln Gln
                1380                1385                1390

CTG CCC CGC TCC ATC TGC AGC CTG CCC TGC CAA CCG GGT GAG CGG AAG      1760
Leu Pro Arg Ser Ile Cys Ser Leu Pro Cys Gln Pro Gly Glu Arg Lys
            1395                1400                1405

AAG ACA GTG AAG GGC ATG CCT TGC TGC TGG CAC TGC GAG CCT TGC ACA      1808
Lys Thr Val Lys Gly Met Pro Cys Cys Trp His Cys Glu Pro Cys Thr
        1410                1415                1420

GGG TAC CAG TAC CAG GTG GAC CGC TAC ACC TGT AAG ACG TGT CCC TAT      1856
Gly Tyr Gln Tyr Gln Val Asp Arg Tyr Thr Cys Lys Thr Cys Pro Tyr
    1425                1430                1435

GAC ATG CGG CCC ACA GAG AAC CGC ACG GGC TGC CGG CCC ATC CCC ATC      1904
Asp Met Arg Pro Thr Glu Asn Arg Thr Gly Cys Arg Pro Ile Pro Ile
1440                1445                1450                1455

ATC AAG CTT GAG TGG GGC TCG CCC TGG GCC GTG CTG CCC CTC TTC CTG      1952
Ile Lys Leu Glu Trp Gly Ser Pro Trp Ala Val Leu Pro Leu Phe Leu
                1460                1465                1470

GCC GTG GTG GGC ATC GCT GCC ACG TTG TTC GTG GTG ATC ACC TTT GTG      2000
Ala Val Val Gly Ile Ala Ala Thr Leu Phe Val Val Ile Thr Phe Val
            1475                1480                1485

CGC TAC AAC GAC ACG CCC ATC GTC AAG GCC TCG GGC CGT GAA CTG AGC      2048
Arg Tyr Asn Asp Thr Pro Ile Val Lys Ala Ser Gly Arg Glu Leu Ser
        1490                1495                1500

TAC GTG CTG CTG GCA GGC ATC TTC CTG TGC TAT GCC ACC ACC TTC CTC      2096
Tyr Val Leu Leu Ala Gly Ile Phe Leu Cys Tyr Ala Thr Thr Phe Leu
    1505                1510                1515

ATG ATC GCT GAG CCC GAC CTT GGC ACC TGC TCG CTG CGC CGA ATC TTC      2144
Met Ile Ala Glu Pro Asp Leu Gly Thr Cys Ser Leu Arg Arg Ile Phe
1520                1525                1530                1535

CTG GGA CTA GGG ATG AGC ATC AGC TAT GCA GCC CTG CTC ACC AAG ACC      2192
Leu Gly Leu Gly Met Ser Ile Ser Tyr Ala Ala Leu Leu Thr Lys Thr
                1540                1545                1550

AAC CGC ATC TAC CGC ATC TTC GAG CAG GGC AAG CGC TCG GTC AGT GCC      2240
Asn Arg Ile Tyr Arg Ile Phe Glu Gln Gly Lys Arg Ser Val Ser Ala
            1555                1560                1565

CCA CGC TTC ATC AGC CCC GCC TCA CAG CTG GCC ATC ACC TTC AGC CTC      2288
Pro Arg Phe Ile Ser Pro Ala Ser Gln Leu Ala Ile Thr Phe Ser Leu
        1570                1575                1580

ATC TCG CTG CAG CTG CTG GGC ATC TGT GTG TGG TTT GTG GTG GAC CCC      2336
Ile Ser Leu Gln Leu Leu Gly Ile Cys Val Trp Phe Val Val Asp Pro
```

```
    1585                1590                1595

TCC CAC TCG GTG GTG GAC TTC CAG GAC CAG CGG ACA CTC GAC CCC CGC      2384
Ser His Ser Val Val Asp Phe Gln Asp Gln Arg Thr Leu Asp Pro Arg
1600                1605                1610                1615

TTC GCC AGG GGT GTG CTC AAG TGT GAC ATC TCG GAC CTG TCG CTC ATC      2432
Phe Ala Arg Gly Val Leu Lys Cys Asp Ile Ser Asp Leu Ser Leu Ile
                1620                1625                1630

TGC CTG CTG GGC TAC AGC ATG CTG CTC ATG GTC ACG TGC ACC GTG TAT      2480
Cys Leu Leu Gly Tyr Ser Met Leu Leu Met Val Thr Cys Thr Val Tyr
            1635                1640                1645

GCC ATC AAG ACA CGC GGC GTG CCC GAG ACC TTC AAT GAG GCC AAG CCC      2528
Ala Ile Lys Thr Arg Gly Val Pro Glu Thr Phe Asn Glu Ala Lys Pro
        1650                1655                1660

ATT GGC TTC ACC ATG TAC ACC ACT TGC ATC GTC TGG CTG GCC TTC ATC      2576
Ile Gly Phe Thr Met Tyr Thr Thr Cys Ile Val Trp Leu Ala Phe Ile
    1665                1670                1675

CCC ATC TTC TTT GGC ACC TCG CAG TCG GCC GAC AAG CTG TAC ATC CAG      2624
Pro Ile Phe Phe Gly Thr Ser Gln Ser Ala Asp Lys Leu Tyr Ile Gln
1680                1685                1690                1695

ACG ACG ACG CTG ACG GTC TCG GTG AGT CTG AGC GCC TCG GTG TCC CTG      2672
Thr Thr Thr Leu Thr Val Ser Val Ser Leu Ser Ala Ser Val Ser Leu
                1700                1705                1710

GGA ATG CTC TAC ATG CCC AAA GTC TAC ATC ATC CTC TTC CAC CCG GAG      2720
Gly Met Leu Tyr Met Pro Lys Val Tyr Ile Ile Leu Phe His Pro Glu
            1715                1720                1725

CAG AAC GTG CCC AAG CGC AAG CGC AGC CTC AAA GCC GTC GTT ACG GCG      2768
Gln Asn Val Pro Lys Arg Lys Arg Ser Leu Lys Ala Val Val Thr Ala
        1730                1735                1740

GCC ACC ATG TCC AAC AAG TTC ACG CAG AAG GGC AAC TTC CGG CCC AAC      2816
Ala Thr Met Ser Asn Lys Phe Thr Gln Lys Gly Asn Phe Arg Pro Asn
    1745                1750                1755

GGA GAG GCC AAG TCT GAG CTC TGC GAG AAC CTT GAG GCC CCA GCG CTG      2864
Gly Glu Ala Lys Ser Glu Leu Cys Glu Asn Leu Glu Ala Pro Ala Leu
1760                1765                1770                1775

GCC ACC AAA CAG ACT TAC GTC ACT TAC ACC AAC CAT GCA ATC              2906
Ala Thr Lys Gln Thr Tyr Val Thr Tyr Thr Asn His Ala Ile
                1780                1785

TAGCGAGTCC ATGGAGCTGA GCAGCAGGAG GAGGAGCCGT GACCCTGTGG AAGGTGCGTC    2966

GGGCCAGGGC CACACCCAAG GGCCCAGCTG TCTTGCCTGC CCGTGGGCAC CCACGGACGT    3026

GGCTTGGTGC TGAGGATAGC AGAGCCCCCA GCCATCACTG CTGGCAGCCT GGGCAAACCG    3086
```

GGTGAGCAAC AGGAGGACGA GGGGCCGGGG CGGTGCCAGG CTACCACAAG AACCTGCGTC 3146

TTGGACCATT GCCCCTCCCG GCCCCAAACC ACAGGGGCTC AGGTCGTGTG GGCCCCAGTG 3206

CTAGATCTCT CCCTCCCTTC GTCTCTGTCT GTGCTGTTGG CGACCCCTCT GTCTGTCTCC 3266

AGCCCTGTCT TTCTGTTCTC TTATCTCTTT GTTTCACCTT TTCCCTCTCT GGCGTCCCCG 3326

GCTGCTTGTA CTCTTGGCCT TTTCTGTGTC TCCTTTCTGG CTCTTGCCTC CGCCTCTCTC 3386

TCTCATCCTC TTTGTCCTCA GCTCCTCCTG CTTTCTTGGG TCCCACCAGT GTCACTTTTC 3446

TGCCGTTTTC TTTCCTGTTC TCCTCTGCTT CATTCTCGTC CAGCCATTGC TCCCCTCTCC 3506

CTGCCACCCT TCCCCAGTTC ACCAAACCTT ACATGTTGCA AAAGAGAAAA AAGGAAAAAA 3566

AATCAAAACA CAAAAAAGCC AAAACGAAAA CAAATCTCGA GTGTGTTGCC AAGTGCTGCG 3626

TCCTCCTGGT GGCCTCTGTG TGTGTCCCTG TGGCCCGCAG CCTGCCCGCC TGCCCCGCCC 3686

ATCTGCCGTG TGTCTTGCCC GCCTGCCCCG CCCGTCTGCC GTCTGTCTTG CCCGCCTGCC 3746

CGCCTGCCCC TCCTGCCGAC CACACGGAGT TCAGTGCCTG GGTGTTTGGT GATGGTTATT 3806

GACGACAATG TGTAGCGCAT GATTGTTTTT ATACCAAGAA CATTTCTAAT AAAAATAAAC 3866

ACATGGTTTT GCAAAAAA 3884

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 912 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Met Pro Gly Lys Arg Gly Leu Gly Trp Trp Trp Ala Arg Leu Pro Leu
 1               5                  10                  15

Cys Leu Leu Leu Ser Leu Tyr Gly Pro Trp Met Pro Ser Ser Leu Gly
            20                  25                  30

Lys Pro Lys Gly His Pro His Met Asn Ser Ile Arg Ile Asp Gly Asp
        35                  40                  45

Ile Thr Leu Gly Gly Leu Phe Pro Val His Gly Arg Gly Ser Glu Gly
    50                  55                  60

Lys Pro Cys Gly Glu Leu Lys Lys Glu Lys Gly Ile His Arg Leu Glu
65                  70                  75                  80

Ala Met Leu Phe Ala Leu Asp Arg Ile Asn Asn Asp Pro Asp Leu Leu
                85                  90                  95
```

Pro Asn Ile Thr Leu Gly Ala Arg Ile Leu Asp Thr Cys Ser Arg Asp
100 105 110

Thr His Ala Leu Glu Gln Ser Leu Thr Phe Val Gln Ala Leu Ile Glu
115 120 125

Lys Asp Gly Thr Glu Val Arg Cys Gly Ser Gly Gly Pro Pro Ile Ile
130 135 140

Thr Lys Pro Glu Arg Val Val Gly Val Ile Gly Ala Ser Gly Ser Ser
145 150 155 160

Val Ser Ile Met Val Ala Asn Ile Leu Arg Leu Phe Lys Ile Pro Gln
165 170 175

Ile Ser Tyr Ala Ser Thr Ala Pro Asp Leu Ser Asp Asn Ser Arg Tyr
180 185 190

Asp Phe Phe Ser Arg Val Val Pro Ser Asp Thr Tyr Gln Ala Gln Ala
195 200 205

Met Val Asp Ile Val Arg Ala Leu Lys Trp Asn Tyr Val Ser Thr Val
210 215 220

Ala Ser Glu Gly Ser Tyr Gly Glu Ser Gly Val Glu Ala Phe Ile Gln
225 230 235 240

Lys Ser Arg Glu Asp Gly Gly Val Cys Ile Ala Gln Ser Val Lys Ile
245 250 255

Pro Arg Glu Pro Lys Ala Gly Glu Phe Asp Lys Ile Ile Arg Arg Leu
260 265 270

Leu Glu Thr Ser Asn Ala Arg Ala Val Ile Ile Phe Ala Asn Glu Asp
275 280 285

Asp Ile Arg Arg Val Leu Glu Ala Ala Arg Arg Ala Asn Gln Thr Gly
290 295 300

His Phe Phe Trp Met Gly Ser Asp Ser Trp Gly Ser Lys Ile Ala Pro
305 310 315 320

Val Leu His Leu Glu Glu Val Ala Glu Gly Ala Val Thr Ile Leu Pro
325 330 335

Lys Arg Met Ser Val Arg Gly Phe Asp Arg Tyr Phe Ser Ser Arg Thr
340 345 350

Leu Asp Asn Asn Arg Arg Asn Ile Trp Phe Ala Glu Phe Trp Glu Asp
355 360 365

Asn Phe His Cys Lys Leu Ser Arg His Ala Leu Lys Lys Gly Ser His
370 375 380

Val Lys Lys Cys Thr Asn Arg Glu Arg Ile Gly Gln Asp Ser Ala Tyr
385 390 395 400

Glu Gln Glu Gly Lys Val Gln Phe Val Ile Asp Ala Val Tyr Ala Met
405 410 415

Gly His Ala Leu His Ala Met His Arg Asp Leu Cys Pro Gly Arg Val
420 425 430

Gly Leu Cys Pro Arg Met Asp Pro Val Asp Gly Thr Gln Leu Leu Lys
435 440 445

Tyr Ile Arg Asn Val Asn Phe Ser Gly Ile Ala Gly Asn Pro Val Thr
450 455 460

Phe Asn Glu Asn Gly Asp Ala Pro Gly Arg Tyr Asp Ile Tyr Gln Tyr
465 470 475 480

Gln Leu Arg Asn Asp Ser Ala Glu Tyr Lys Val Ile Gly Ser Trp Thr
485 490 495

Asp His Leu His Leu Arg Ile Glu Arg Met His Trp Pro Gly Ser Gly
500 505 510

Gln Gln Leu Pro Arg Ser Ile Cys Ser Leu Pro Cys Gln Pro Gly Glu
515 520 525

Arg Lys Lys Thr Val Lys Gly Met Pro Cys Cys Trp His Cys Glu Pro
530 535 540

Cys Thr Gly Tyr Gln Tyr Gln Val Asp Arg Tyr Thr Cys Lys Thr Cys
545 550 555 560

Pro Tyr Asp Met Arg Pro Thr Glu Asn Arg Thr Gly Cys Arg Pro Ile
565 570 575

Pro Ile Ile Lys Leu Glu Trp Gly Ser Pro Trp Ala Val Leu Pro Leu
580 585 590

Phe Leu Ala Val Val Gly Ile Ala Ala Thr Leu Phe Val Val Ile Thr
595 600 605

Phe Val Arg Tyr Asn Asp Thr Pro Ile Val Lys Ala Ser Gly Arg Glu
610 615 620

Leu Ser Tyr Val Leu Leu Ala Gly Ile Phe Leu Cys Tyr Ala Thr Thr
625 630 635 640

Phe Leu Met Ile Ala Glu Pro Asp Leu Gly Thr Cys Ser Leu Arg Arg
645 650 655

Ile Phe Leu Gly Leu Gly Met Ser Ile Ser Tyr Ala Ala Leu Leu Thr
660 665 670

Lys Thr Asn Arg Ile Tyr Arg Ile Phe Glu Gln Gly Lys Arg Ser Val
675 680 685

Ser Ala Pro Arg Phe Ile Ser Pro Ala Ser Gln Leu Ala Ile Thr Phe
690 695 700

Ser Leu Ile Ser Leu Gln Leu Leu Gly Ile Cys Val Trp Phe Val Val
705 710 715 720

Asp Pro Ser His Ser Val Val Asp Phe Gln Asp Gln Arg Thr Leu Asp
725 730 735

Pro Arg Phe Ala Arg Gly Val Leu Lys Cys Asp Ile Ser Asp Leu Ser
740 745 750

Leu Ile Cys Leu Leu Gly Tyr Ser Met Leu Leu Met Val Thr Cys Thr
755 760 765

Val Tyr Ala Ile Lys Thr Arg Gly Val Pro Glu Thr Phe Asn Glu Ala
770 775 780

Lys Pro Ile Gly Phe Thr Met Tyr Thr Thr Cys Ile Val Trp Leu Ala
785 790 795 800

Phe Ile Pro Ile Phe Phe Gly Thr Ser Gln Ser Ala Asp Lys Leu Tyr
805 810 815

Ile Gln Thr Thr Thr Leu Thr Val Ser Val Ser Leu Ser Ala Ser Val
820 825 830

Ser Leu Gly Met Leu Tyr Met Pro Lys Val Tyr Ile Ile Leu Phe His
835 840 845

Pro Glu Gln Asn Val Pro Lys Arg Lys Arg Ser Leu Lys Ala Val Val
850 855 860

Thr Ala Ala Thr Met Ser Asn Lys Phe Thr Gln Lys Gly Asn Phe Arg
865 870 875 880

Pro Asn Gly Glu Ala Lys Ser Glu Leu Cys Glu Asn Leu Glu Ala Pro
885 890 895

Ala Leu Ala Thr Lys Gln Thr Tyr Val Thr Tyr Thr Asn His Ala Ile
900 905 910

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

ACAACGACAC ACCCGTGGTC AA 22

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

TCCGGTCGGG AGCTCTGCTA 20

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GCTACTCTGC CCTGCTGACC AAGAC 25

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCAATGCGGT TGGTCTTGGT 20

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

AAGTTCATCG GCTTCACCAT GTACAC 26

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GATGCAGGTG GTGTACATGG TGAA 24

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

TCTCCGGTTG GAAGAGGATG ATGT 24

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

CTCTGCCTAC TCCTCAGCCT TTA 23

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GCACAAAGGT CAGTGACTGC TC 22

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

AGCTCGGTCT CCATCAT 17

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TGATGTCATC CTCGTTGGC 19

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AGAGCTCGAG AAACAGGATT CATGAAGATG 30

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GCAATCTAGA ATCACAGAGA TGAGGTG 27

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GGGTGTCTAG AGATTTCCGA G 21

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

AGAGGCGAAG GATGTTG 17

CLAIMS

1. A DNA isolate encoding human mGluR3 or human mGluR4.

2. A DNA isolate according to claim 1 encoding human mGluR3 having an amino acid sequence at least 97% identical to the sequence of SEQ ID NO:2 or human mGluR4 having an amino acid sequence at least 97% identical to the sequence of SEQ ID NO:4.

3. A DNA isolate according to claim 2 encoding human mGluR3 having the amino acid sequence of SEQ ID NO:2 or human mGluR4 having the amino acid sequence of SEQ ID NO:4.

4. A DNA isolate according to claim 1 containing a sequence at least 90% identical to the mGluR3 coding sequence set forth from position 253 or 259 to position 2889 of SEQ ID NO:1 or a sequence at least 90% identical to the mGluR4 coding sequence set forth from position 171 to position 2906 of SEQ ID NO:3.

5. A DNA isolate according to claim 1 containing a nucleotide sequence at least 90% identical to the sequence set forth in SEQ ID NO:1 or a nucleotide sequence at least 90% identical to the sequence of SEQ ID NO:3.

6. A DNA isolate according to claim 4 containing the mGluR3 coding sequence set forth from position 253 or 259 to position 2889 of SEQ ID NO:1 or the mGluR4 coding sequence set forth from position 171 to position 2906 of SEQ ID NO:3.

7. A DNA isolate according to claim 5 containing the nucleotide sequence set forth in SEQ ID NO:1 or the nucleotide sequence set forth in SEQ ID NO:3.

8. Human mGluR3 or human mGluR4.

9. Human mGluR3 according to claim 8 having an amino acid sequence at least 97% identical to the sequence of SEQ ID NO:2 or human mGluR4 according to claim 8 having an amino acid sequence at least 97% identical to the sequence of SEQ ID NO:4.

10. Human mGluR3 according to claim 8 having the amino acid sequence of SEQ ID NO:2 or human mGluR4 according to claim 8 having the amino acid sequence of SEQ ID NO:4.

11. An oligonucleotide fragment which is at least 12 nucleotides in length and which has the sequence of a portion of the sequence of the DNA isolate as claimed in claim 7.

12. An oligonucleotide probe or an oligonucleotide primer which hybridizes to the DNA isolate as claimed in claim 7.

13. A polypeptide fragment of human mGluR3 or human mGluR4 as claimed in claim 10, which fragment is at least 8 amino acids in length.

14. A recombinant DNA vector which comprises a DNA isolate as claimed in any one of claims 1 to 7.

15. A host cell transformed or transfected with a vector as claimed in claim 14.

16. A method for producing a host cell expressing human mGluR3 or human mGluR4 as claimed in any one of claims 8 to 10, which method comprises

(a) transforming or transfecting cells with a vector as claimed in claim 14,

(b) culturing the cells under conditions in which the mGluR3 or mGluR4 is expressed; and

(c) recovering a host cell expressing the mGluR3 or mGluR4 from the culture.

17. A method of identifying or quantitively determining

in a sample a target nucleic acid encoding human mGluR3 or human mGluR4 as claimed in any one of claims 8 to 10, which method comprises

(a) contacting a probe as claimed in claim 12 with the sample; and

(b) detecting or quantitatively determining hybridization of the probe with any target nucleic acid present in the sample.

18. An antibody specific for human mGluR3 or human mGluR4 as claimed in any one of claims 8 to 10.

19. A method for identifying a compound which binds to human mGluR3 or human mGluR4 receptor as claimed in any one of claims 8 to 10, which method comprises

(a) contacting the compound with the receptor, and

(b) measuring any binding of the compound to the receptor.

20. A method according to claim 19 wherein the compound is contacted with whole cells and/or cell membranes containing the receptor.

21. A method according to claim 19 or 20 wherein the binding of the compound to the receptor is compared with the binding to the receptor of a control ligand that is known to bind to the receptor.

22. A method for identifying a compound which modulates the activity of human mGluR3 or human mGluR4 receptor as claimed in any one of claims 8 to 10, which method comprises

(a) contacting the compound with the receptor, and

(b) measuring any change in second messenger receptor activity.

23. A method according to claim 22 wherein the compound

is contacted with whole cells containing the receptor.

24. A method according to claim 22 or 23 wherein any change in basal or stimulated adenylate cyclase activity is measured.

25. A compound identified by a method as claimed in any one of claims 19 to 24.

26. A combination of human mGluR3 or human mGluR4 receptor as claimed in any one of claims 8 to 10 and an agent which modulates the second messenger activity of the receptor.

27. A method of modulating the second messenger activity of human mGluR3 or human mGluR4 receptor as claimed in any one of claims 8 to 10, which method comprises contacting the receptor with a compound which modulates the second messenger activity of the receptor.

```
  1 MLTRLQVLTLALFSKGFLLSGLDHNFLRREIKIEGDLVLGGLFPINEKGT
          I M                 M

 51 GTEECGRINEDRGIQRLEAMLFAIDEINKDDYLLPGVKLGVHILDTCSRD
                                  N

101 TYALEQSLEFVRASLTKVDEAEYMCPDGSYAIQENIPLLIAGVIGGSYSS

151 VSIQVANLLRLFQIPQISYASTSAKLSDKRRYDYFARTVPPDFYQAKAMA

201 EILRFFNWTYVSTVASEGDYGETGIEAFEQEARLRNICIATAEKVGRSNI

251 RKSYDSVIRELLQKPNARVVVLFMRSDDSRELIAAASRANASFTWVASDG
                                        N V

301 WGAQESIIKGSEHVAYGAITLELASQPVRQFDRYFQSLNPYNNHRNPWFR
           V                 H

351 DFWEQKFQCSLQNKRNHRRVCDKHLAIDSSNYEQESKIMFVVNAVYAMAH
                      Q

401 ALHKMQRTLCPNTTKLCDAMKILDGKKLYKDYLLKINFTAPFNPNKDADS
                                  E               G

451 IVKFDTFGDGMGRYNVFNFQNVGGKYSYLKVGHWAETLSLCVNSIHWSRN
                      L QT                    D

501 SVPTSQCSDPCAPNEMKNMQPGDVCCWICIPCEPYEYLADEFTCMDCGSG
                                          V         P
                                     TMD-I
551 QWPTADLTGCYDLPEDYIRWEDAWAIGPVTIACLGFMCTCMVVTVFIKHN
           S   N      K                 L   I I
                        TMD-II
601 NTPLVKASGRELCYILLFGVGLSYCMTFFFIAKPSPVICALRRLGLGSSF
                        S                          T
     TMD-III                                  TMD-IV
651 AICYSALLTKTNCIARIFDGVKNGAQRPKFISPSSQVFICLGLILVQIVM
                                              TMD-V
701 VSVWLILEAPGTRRYTLAEKRETVILKCNVKDSSMLISLTYDVILVILCT
            T        P                         V
                               TMD-VI
751 VYAFKTRKCPENFNEAKFIGFTMYTTCIIWLAFLPIFYVTSSDYRVQTTT
              TMD-VII
801 MCISVSLSGFVVLGCLFAPKVHIILFQPQKNVVTHRLHLNRFSVSGTGTT
                           V                       A

851 YSQSSASTYVPTVCNGREVLDSTTSSL
```

Fig. 1

```
  1 MPGKRGLGWWWARLPLCLLLSLYGPWMPSSLGKPKGHPHMNSIRIDGDIT
     S   G WA                    A  V

 51 LGGLFPVHGRGSEGKPCGELKKEKGIHRLEAMLFALDRINNDPDLLPNIT
                        A

101 LGARILDTCSRDTHALEQSLTFVQALIEKDGTEVRCGSGGPPIITKPERV

151 VGVIGASGSSVSIMVANILRLFKIPQISYASTAPDLSDNSRYDFFSRVVP

201 SDTYQAQAMVDIVRALKWNYVSTVASEGSYGESGVEAFIQKSREDGGVCI
                              L                       N

251 AQSVKIPREPKAGEFDKIIRRLLETSNARAVIIFANEDDIRRVLEAARRA
                   T       K         GI

301 NQTGHFFWMGSDSWGSKIAPVLHLEEVAEGAVTILPKRMSVRGFDRYFSS
                            S    R

351 RTLDNNRRNIWFAEFWEDNFHCKLSRHALKKGSHVKKCTNRERIGQDSAY
                                                  I

401 EQEGKVQFVIDAVYAMGHALHAMHRDLCPGRVGLCPRMDPVDGTQLLKYI

451 RNVNFSGIAGNPVTFNENGDAPGRYDIYQYQLRNDSAEYKVIGSWTDHLH
                                                  G

501 LRIERMHWPGSGQQLPRSICSLPCQPGERKKTVKGMPCCWHCEPCTGYQY
          Q                                       A

                                                    TMD-I
551 QVDRYTCKTCYPDMRPTENRTGCRPIPIIKLEWGSPWAVLPLFLAVVGIA
                                S Q    V    D

                                                TMD-II
601 ATLFVVITFVRYNDTPIVKASGRELSYVLLAGIFLCYATTFLMIAEPDLG
           V

                        TMD-III
651 TCSLRRIFLGLGMSISYAALLTKTNRIYRIFEQGKRSVSAPRFISPASQL

               TMD-IV
701 AITFSLISLQLLGICVWFVVDPSHSVVDFQDQRTLDPRFARGVLKCDISD
        I

                  TMD-V                                 TMD-VI
751 LSLICLLGYSMLLMVTCTVYAIKTRGVPETFNEAKPIGFTMYTTCIVWLA

                                                TMD-VII
801 FIPIFFGTSQSADKLYIQTTTLTVSVSLSASVSLGMLYMPKVYIILFHPE

851 QNVPKRKRSLKAVVTAATMSNKFTQKGNFRPNGEAKSELCENLEAPALAT
                                                    T

901 KQTYVTYTNHAI
```

Fig. 2

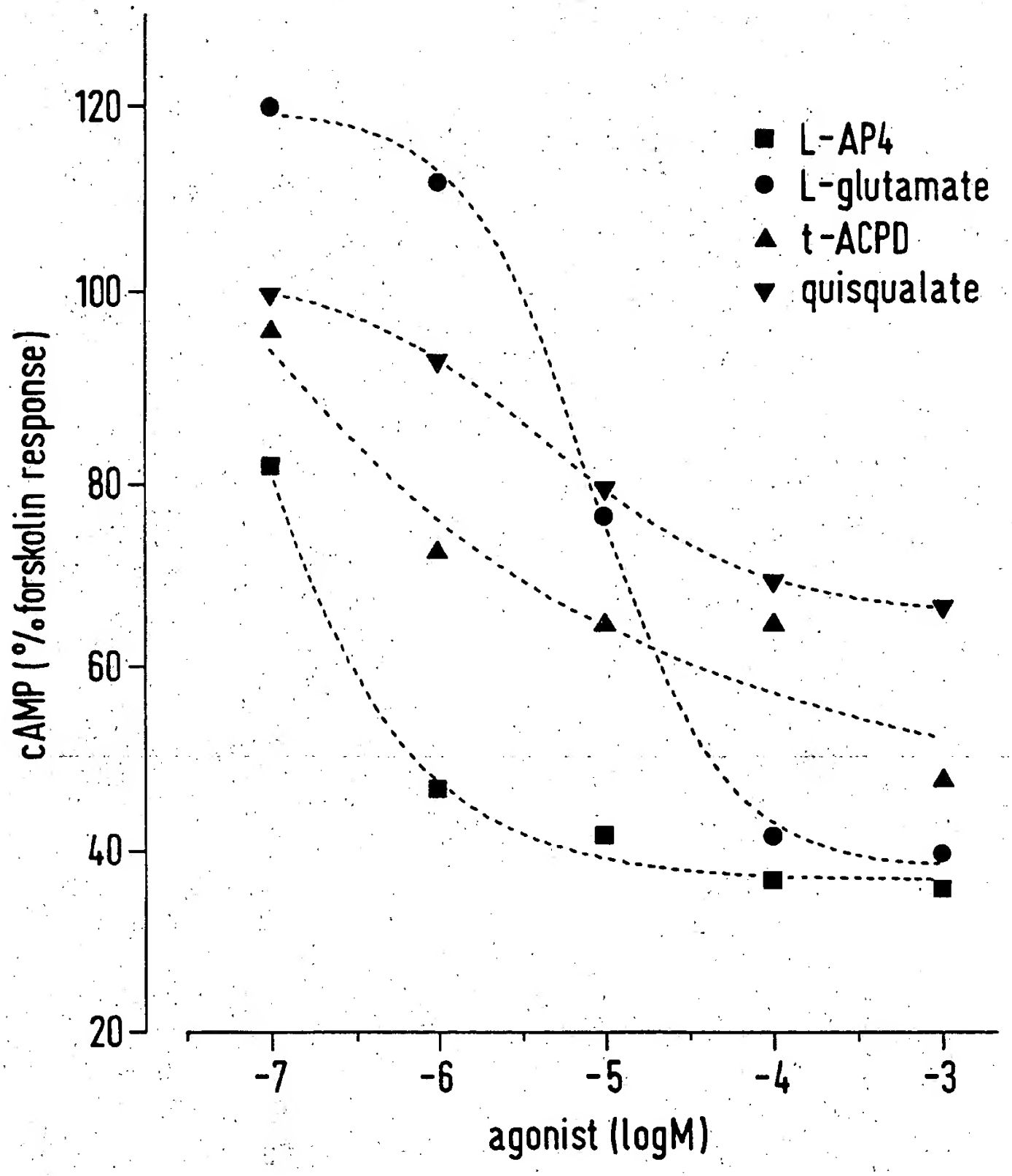
120
100
80
60
40
20
cAMP (% forskolin response)
-7
-6
-5
-4
-3
agonist (logM)
L-AP4
L-glutamate
t-ACPD
quisqualate

Fig.3